Audrey Terras

Harmonic Analysis on Symmetric Spaces—Euclidean Space, the Sphere, and the Poincaré Upper Half-Plane

Second Edition

Springer

Harmonic Analysis on Symmetric Spaces—Euclidean Space, the Sphere, and the Poincaré Upper Half-Plane

Audrey Terras

Harmonic Analysis on Symmetric Spaces—Euclidean Space, the Sphere, and the Poincaré Upper Half-Plane

Second Edition

Springer

Audrey Terras
Department of Mathematics
University of California at San Diego
La Jolla, CA, USA

ISBN 978-1-4614-7971-0 ISBN 978-1-4614-7972-7 (eBook)
DOI 10.1007/978-1-4614-7972-7
Springer New York Heidelberg Dordrecht London

Library of Congress Control Number: 2013939730

Mathematics Subject Classification: 42XX, 22-XX, 43A85, 11F72, 11Fxx, 30F35, 33B15, 33C55, 34B24, 35Pxx, 44A10, 45Bxx, 46Fxx, 53A35, 58C40, 34L40, 35J10, 81Qxx, 58D19, 60F05, 47B15, 11T71, 11T60, 28C10, 05Cxx, 94B05, 52C23

© Springer Science+Business Media New York 1985, 2013
This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Introductory Quotation

I have myself always thought of a mathematician as in the first instance an *observer*, a man who gazes at a distant range of mountains and notes down his observations. His object is simply to distinguish clearly and notify to others as many different peaks as he can .... But when he sees a peak he believes that it is there simply because he sees it. If he wishes someone else to see it, he points to it, either directly or through the chain of summits which led him to recognize it himself. When his pupil also sees it, the research, the argument, the *proof* is finished.

The analogy is a rough one, but I am sure that it is not altogether misleading. If we were to push it to its extreme we should be led to a rather paradoxical conclusion; that there is, strictly, no such thing as mathematical proof; that we can, in the last analysis, do nothing but *point*; that proofs are what Littlewood and I call *gas*, rhetorical flourishes designed to affect psychology, pictures on the board in the lecture, devices to stimulate the imagination of pupils. This is plainly not the whole truth, but there is a good deal in it.

—From Hardy [250, Vol. 7, p. 598].

To all women mathematicians,
to my parents,
to my POSSLQ.

Preface to the First Edition

Since its beginnings with Fourier (and as far back as the Babylonian astronomers), harmonic analysis has been developed with the goal of unraveling the mysteries of the physical world of quasars, brain tumors, and so forth, as well as the mysteries of the nonphysical, but no less concrete, world of prime numbers, diophantine equations, and zeta functions. Quoting Courant and Hilbert, in the preface to the first German edition of *Methods of Mathematical Physics*: “Recent trends and fashions have, however, weakened the connection between mathematics and physics.” Such trends are still in evidence, harmful though they may be. My main motivation in writing these notes has been a desire to counteract this tendency towards specialization and describe applications of harmonic analysis in such diverse areas as number theory (which happens to be my specialty), statistics, medicine, geophysics, and quantum physics. I remember being quite surprised to learn that the subject is useful. My graduate education was that of the 1960s. The standard mathematics graduate course proceeded from Definition 1.1.1 to Corollary 14.5.59, with no room in between for applications, motivation, history, or references to related work. My aim has been to write a set of notes for a very different sort of course.

A second impulse pushing me toward the typewriter was the knowledge that in the past 30 years there have been some really exciting discoveries in the field of harmonic analysis on symmetric spaces and their fundamental domains for discrete isometry groups—the work of Harish-Chandra, Helgason, Langlands, Maass, Selberg, and others. It is time that these ideas received an exposition comprehensible to the average applied mathematician, number theorist, etc. In particular, I believe that many of the results to be described have interesting implications for statistical physics and number theory.

The outline of the book can be sketched as follows. Chapter 1 concerns Euclidean Fourier analysis and its applications to the solution of the wave and heat equations, the study of potential functions of crystals, as well as zeta functions of algebraic number fields, for example. Chapter 2 deals with spherical Fourier analysis and its connections with the Euclidean theory. There are applications to CAT scanners, the solar corona, and the Zeeman effect for the hydrogen atom in a magnetic field.

Chapter 3 studies non-Euclidean Fourier analysis on the Poincaré or Lobatchevsky upper half-plane H, with an elementary discussion of the work of Harish-Chandra and Helgason in this special case. The main idea is to use the group invariance under $SL(2,\mathbb{R})$—the special linear group of all 2×2 real matrices of determinant one—to determine the spectral measure in the non-Euclidean Fourier inversion formula via the asymptotics of the special functions involved. Applications include the solution of the Dirichlet problem for cones, wedges, and other domains in Euclidean space, as was discovered by Mehler, Fock, and Lebedev long ago. One can also solve the non-Euclidean heat and wave equations on H itself. These results will be used to obtain a non-Euclidean central limit theorem with applications to the statistics of transmission lines. The results of Chap. 3 have different interpretations when one realizes that the group $SL(2,\mathbb{R})$ can be identified with many other Lie groups; e.g., $Sp(1,\mathbb{R})$ the group of 2×2 real symplectic matrices g such that ${}^t gJg = J$ if $J = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$, or the group $SU(1,1)$ of 2×2 complex matrices g such that ${}^t\overline{g}Mg = M$ if $M = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$. Finally, $SL(2,\mathbb{R})$ is locally isomorphic to the Lorentz-type group $SO(2,1)$ of rea1 3×3 matrices of determinant one preserving the quadratic form $x_1^2 + x_2^2 - x_3^2$. However, the higher-dimensional analogues of these groups are distinct.

Non-Euclidean analogues of Fourier series also make their appearance in Chap. 3. This is the Fourier inversion formula for functions on H that are periodic under the group $SL(2,\mathbb{Z})$ acting via fractional linear transformation with integer entries and determinant one. The methods of the first part of the chapter work in this case to find the spectral measure on the continuous part of the spectrum of the non-Euclidean Laplacian on $H/SL(2,\mathbb{Z})$ just as they did for H itself. However, there is, in addition, a discrete part of the spectrum, which remains as mysterious as the quanta in quantum mechanics. Applications in this section include the solution of the non-Euclidean heat equation on the fundamental domain $H/SL(2,\mathbb{Z})$, the computation of class numbers of imaginary quadratic fields, and Peter Sarnak's use of the Selberg trace formula to say something about the asymptotics of units in real quadratic fields. Recently Hurt [308, 309] has described applications of the trace formula in quantum-statistical mechanics. Chapter 1 of Volume II, i.e., [667], will include another application—Kaori Imai Ota's extension of the converse result in Hecke's theory of the relation between Dirichlet series such as Riemann's zeta function, and modular forms such as the theta function. Imai Ota extends the theory to Siegel modular forms of genus 2. Examples of such modular forms are the theta functions appearing in the study of abelian integrals. These theta functions arise in recent work on the Korteweg–DeVries equation, as well as in Sonya Kovalevsky's solution of the third known case of the problem of the motion of a rigid body about a fixed point.

Volume II, Chap. 1 will extend harmonic analysis to the symmetric space $\mathcal{P}_n$ of positive definite real $n \times n$ matrices Y as well as to the Minkowski fundamental domain for $\mathcal{P}_n/GL(n,\mathbb{Z})$, where $GL(n,\mathbb{Z})$ is the modular group of $n \times n$ integer matrices of determinant ± 1 (with the action Y goes to ${}^tAYA$, if ${}^tA =$ transpose of A).

This is the prototype of the general theory for the symmetric space of a noncompact semisimple (or reductive) Lie group. Applications of Volume II, Chap. 1 include the study of the Wishart distribution in statistics, integrals arising in statistical quantum mechanics, lattice packings of spheres and Hilbert's 18th problem, and integral tests for convergence of sums over matrix variables. The theory that underlies the development of the higher-dimensional Selberg trace formula will also be discussed. The most important part concerns various matrix argument analogues of certain useful special functions such as spherical, gamma, Bessel, Whittaker, zeta, and L-functions. There are many thesis topics and intriguing open questions here, some of which are quite basic. Thus Volume II has a somewhat unfinished aspect.

Volume II, Chap. 2 is much more crude. Perhaps it should be considered merely as a guide to the literature. My goal was to present an introduction to some of the work of Maass and Siegel, because this work had motivated much of the development in Volume II, Chap. 1 (and even some of Volume I, Chap. 3). Another stimulus was the hope that by fitting various examples into the general picture, one would derive a clearer understanding of the examples. There are also quite interesting relations between the various symmetric spaces and fundamental domains, some of which have been of great use in number theory. Here the key word is "base change."

I would like to thank those people who have discussed these notes with me over the years. It is becoming hopeless to name them all. I would also like to note my indebtedness to certain books and papers that influenced me greatly during the time I began to write this manuscript: Courant and Hilbert [111], Dym and McKean [147], Elstrodt [156], Gangolli [191, 192], Hejhal [261, 262], Helgason [276], Kubota [375], Lang [388], Lebedev [401], Maass [437, 438], Mennicke [463], Minkowski [471], Selberg [569], and Siegel [597, 599, 600].

I feel now that I am probably quite far from reaching the goals that motivated my writing. Certainly the average engineer will not want to think about the intricacies of $GL(n,\mathbb{Z})$ presented in Volume II, without some more convincing applications. Number theorists are just beginning to see the utility of that chapter. But I have to admit that the chapter is still in quite a preliminary state. The final word cannot yet be said. For this reason (not to mention the size of each volume), I have decided to split these notes into two volumes.

I still hope that both volumes will be useful to a beginner in the subject. The book has been used in parts of mathematics graduate courses at U.C.S.D. and M.I.T. since 1978. These courses had names like Lie groups, harmonic analysis, mathematical physics, and number theory. The last revision was made at the Institute for Advanced Study, Princeton, during the first months of the year of Orwell.

These notes do assume some things of the reader. At the beginning, advanced calculus (with a little measure theory) suffices, along with the ability to look at a partial differential equation or a Bessel function without flinching. By Volume II, more is required. The reader has to look at a matrix argument Bessel function without flinching. [Throughout these volumes I also assume that the reader is familiar with the standard facts of modern applied algebra such as the beginnings of the theory of groups, fields, and rings, the basic facts about the integers $\mathbb{Z}$ and $\mathbb{Z}/n\mathbb{Z}$.]

One final comment: many important details have been consigned to the exercises. This happened principally because the author likes participatory democracy. However, sometimes the author got carried away with the exercises. But I did try to include references to the proofs. Perhaps later [should I live another 100 years] an answer book will appear.

I feel impelled to add a P.S. of warning to the reader. This author is incredibly bad at proofreading. So, if something appears weird in a formula, feel free to change a letter, insert a missing Fourier transform, correct my spelling or history, etc. I would be very grateful if you would send me a list of the errors that you have found. And I would like to thank those that have done so in the past.

La Jolla, CA Audrey Terras

Preface to the Second Edition

Almost 30 years have passed since Volume I appeared. Much has changed. I am retired. I am happy to have had 25 PhD students. My POSSLQ is now called a partner. I am not using a typewriter. You may read the book as an e-book on a tiny tablet. I have. More women receive PhDs in mathematics in the United States and go on to become full professors at top universities. Obama is president. Yeah!

Some things, however, never change. Basic mathematics is as it ever was. The average person in the United States thinks of a mathematician as someone who can balance his or her checkbook. Publishers, bookstores, libraries, and universities are in trouble. My country is still involved in too many horrible wars.

I tried not to change much in the first edition. I made the corrections I knew about and added some updates on new developments in harmonic analysis on symmetric spaces, keeping myself to those developments that fit in with the spirit of the original. No adeles here! I also added a few figures drawn by Mathematica or Matlab. Again, though I will correct what I know of errors in the first edition, new errors may have been introduced in TeXing the old book. And, yes, I am still bad at proofreading. So reader beware!

Many advances have been made since 1985. We now have Mathematica, Matlab, Sage, Scientific Workplace, Google, Wikipedia. We can post math. papers on the web and people all over the world can read them. No more mailing of poorly typed preprints. But sadly, also, no more handwritten letters. Computers with the power of the supercomputers of yesteryear are now on our desks or even in our purses or jacket pockets. *Star Trek* communicators look clumsy compared with our cell phones.

In the new parts of this second edition, I will discuss wavelets, quasicrystals, modular knots, and also provide a glimpse at the way in which modular forms appeared in the proofs of some of the oldest conjectures in number theory. I will also consider finite analogues of symmetric spaces and their applications in computer science—Ramanujan graphs.

Finally, I thank the makers of Scientific Workplace for making it easier for me to deal with this book.

Encinitas, CA Audrey Terras

Contents

1 Flat Space: Fourier Analysis on $\mathbb{R}^m$ 1

1.1 Distributions or Generalized Functions 1

1.2 Fourier Integrals 8

1.2.1 Fourier Transforms 9

1.2.2 Applications to Partial Differential Equations 17

1.2.3 Laplace Transforms 21

1.2.4 Application to Probability and Statistics: The Central Limit Theorem 25

1.3 Fourier Series and the Poisson Summation Formula 30

1.3.1 Fourier Series 31

1.3.2 Vibrating Drums 36

1.3.3 The Poisson Summation Formula 39

1.3.4 Spectroscopy and the Search for Hidden Periodicities 46

1.3.5 Poisson's Sum Formula as a Trace Formula 53

1.3.6 Schrödinger Eigenvalues 58

1.4 Mellin Transforms, Epstein and Dedekind Zeta Functions 60

1.4.1 Mellin Transforms 60

1.4.2 Epstein's Zeta Function 63

1.4.3 Algebraic Number Theory 70

1.4.4 Crystallography 83

1.5 Finite Symmetric Spaces, Wavelets, Quasicrystals, Weyl's Criterion 88

1.5.1 Symmetric Spaces and Their Finite Analogues 89

1.5.2 Wavelets 94

1.5.3 Quasicrystals 97

1.5.4 Weyl's Criterion for Uniform Distribution 101

2 A Compact Symmetric Space: The Sphere 107

2.1 Fourier Analysis on the Sphere 107

2.1.1 The Sphere as a Symmetric Space 107

2.1.2 Spherical Harmonics 112

2.1.3 Quantum Mechanics: The Hydrogen Atom 118

2.1.4 The Sun's Magnetic Field 121
2.1.5 Connections with Group Representations 125
2.1.6 Integral Equations for Spherical Harmonics 131
2.2 $O(3)$ and $\mathbb{R}^3$: The Radon Transform 133
2.2.1 Harmonic Analysis on $\mathbb{R}^3$ in Spherical Polar Coordinates and Spectral Measures 134
2.2.2 CAT Scanners and the Radon Transform 142

3 The Poincaré Upper Half-Plane 149
3.1 Hyperbolic Geometry 149
3.1.1 The Basics: Arc Length, Area, Laplacian 149
3.1.2 Microwave Engineering: The Smith Chart 156
3.1.3 $SL(\mathbf{2},\mathbb{R})$ as a Lorentz-Type Group 162
3.2 Harmonic Analysis on H 164
3.2.1 Introduction 164
3.2.2 Harmonic Analysis on H in Rectangular Coordinates 165
3.2.3 Harmonic Analysis on H in Geodesic Polar Coordinates 171
3.2.4 The Helgason Transform on H 178
3.2.5 The Heat Equation on H in Rectangular Coordinates 185
3.2.6 The Heat Equation on H Using Helgason's Transform 187
3.2.7 The Central Limit Theorem for K-Invariant Random Variables on H. Transmission Lines with Random Inhomogeneities 189
3.2.8 Some Remarks on the Uses in Theoretical Physics of Harmonic Analysis on Groups Such as $SL(2,\mathbb{R})$ 196
3.3 Fundamental Domains for Discrete Subgroups Γ of $G = SL(2,\mathbb{R})$ 196
3.3.1 Introduction 197
3.3.2 Fundamental Domain for $SL(2,\mathbb{Z})$ 203
3.3.3 Computation of Class Numbers of Imaginary Quadratic Fields 207
3.3.4 Dirichlet or Poincaré Polygon or Normal Fundamental Region 208
3.3.5 Other Discrete Subgroups of $SL(2,\mathbb{R})$ 210
3.3.6 Riemann Surface of the Fundamental Domain 212
3.3.7 Triangle Groups and Quaternion Groups 215
3.3.8 Finite Upper Half-Planes and Their Tessellations 221
3.4 Modular or Automorphic Forms-Classical 227
3.4.1 Definitions, Eisenstein Series 227
3.4.2 The Discriminant, the Weierstrass $\wp$-Function, and Ramanujan's Tau Function 235
3.4.3 The Modular Invariant 240
3.4.4 Theta Functions 242
3.4.5 Elliptic Integrals and Elliptic Curves 248

3.4.6 The Connection Between Theta Functions and Coding Theory 252
3.4.7 Poincaré Series 253
3.4.8 Some Estimates for Fourier Coefficients of Modular Forms 255
3.4.9 The Representation of Integers by Quadratic Forms 256
3.4.10 Korteweg–DeVries Equation 257
3.5 Modular Forms: Not So Classical—Maass Waveforms 258
3.5.1 Maass Waveforms 258
3.5.2 Maass Cusp Forms 268
3.5.3 Computations of Maass Cusp Forms 272
3.5.4 Elementary Estimates of Fourier Coefficients and Eigenvalues 284
3.5.5 Dimensions of Spaces of Maass Cusp Forms 287
3.5.6 Finite Analogues of Eisenstein Series 289
3.6 Modular Forms and Dirichlet Series. Hecke Theory and Generalizations 292
3.6.1 Dirichlet Series Corresponding to Holomorphic Modular Forms: The Hecke Correspondence 292
3.6.2 Dirichlet Series Corresponding to Maass Waveforms 295
3.6.3 Remarks on Extensions to Congruence Subgroups 299
3.6.4 Hecke Operators on Holomorphic Forms 302
3.6.5 Hecke Operators for Maass Wave Forms 306
3.6.6 Rankin–Selberg Method, Distribution of Horocycles in the Fundamental Domain, and Modular Symbols 312
3.7 Harmonic Analysis on the Fundamental Domain 316
3.7.1 Introduction 316
3.7.2 Spectral Resolution of Δ on $SL(2,\mathbb{Z})\backslash H$ 320
3.7.3 A Non-Euclidean Poisson Summation Formula 333
3.7.4 Selberg's Trace Formula for $SL(2,\mathbb{Z})$ 343
3.7.5 Applications of Selberg's Trace Formula 360
3.7.6 Tables Summarizing the Main Results 367
3.7.7 Modular Knots 367
3.7.8 Finite Analogue of the Selberg Trace Formula 374

References 377

Index 403

Chapter 1
Flat Space: Fourier Analysis on $\mathbb{R}^m$

> There was no time to waste. It's possible to grasp alef-null-sized collections once you're in your aethereal body… but you need some to look at. My job right now was to generate infinities.
>
> …
>
> I slipped out of my physical body and began running around the room, I did alef-null laps… and then the whole rest of the class joined in….
>
> "What happened?" a kid with glasses and dandruff asked. "Did you hypnotize us?"
>
> …
>
> "The main thing is that you saw infinity," I said….
>
> —From *White Light*, by Rudy Rucker, Ace Books; NY, 1980, pp. 248–249. Reprinted by permission.

1.1 Distributions or Generalized Functions

> Formalism denies the status of mathematics to most of what has been commonly understood to be mathematics, and can say nothing about its growth. None of the "creative" periods and hardly any of the "critical" periods of mathematical theories would be admitted into the formalist heaven, where mathematical theories dwell like the seraphim, purged of all the impurities of earthly uncertainty. Formalists, though, usually leave open a small back door for fallen angels…. On those terms Newton had to wait four centuries…. Dirac is more fortunate: Schwartz saved his soul during his lifetime.
> —From Lakatos [384, p. 2].

In 1927 P. A. M. Dirac [135] introduced "a function" $\delta(x)$, which he postulated to have the property

$$\int_{-\infty}^{+\infty} f(x)\delta(a-x)dx = f(a) \tag{1.1}$$

if $f(x)$ is continuous near $x = a$. It is easy to see that no such function exists. However, delta has been legalized by Laurent Schwartz and others. And we shall sometimes find it very convenient to have access to this theory of generalized functions or distributions. Thus we begin by summarizing these results. The reader

A. Terras, *Harmonic Analysis on Symmetric Spaces—Euclidean Space, the Sphere, and the Poincaré Upper Half-Plane*, DOI 10.1007/978-1-4614-7972-7_1,
© Springer Science+Business Media New York 2013

who wants more details should consult any of the following references: Yvonne Choquet-Bruhat et al. [87], DeJager [124,125] Dieudonné [134], Friedlander [186], Gelfand and Shilov [206], Horvath [305], Korevaar [371], Maurin [458], Schwartz [565–567], Strichartz [643], Treves [684], and Vladimirov [706]. The concept of distribution is a very natural one in applied mathematics. It can be used to represent an impulse, point mass, or point charge, for example. It also provides a natural method to use in obtaining fundamental solutions of partial differential equations, as we shall see in Sect. 1.2.

There are two definitions of distribution. One of these goes as follows. Let $\mathcal{D}$ be the space of **test functions** $f : \mathbb{R}^m \to \mathbb{C}$ such that f and all partials of all orders of f are continuous and vanish off a bounded set. Then a **distribution** T is a continuous linear functional $T : \mathcal{D} \to \mathbb{C}$. Here we need a topology or notion of convergence of sequences of test functions in order to say what we mean by the continuity of T. We say that **a sequence of test functions converges** if they all vanish off the same bounded set and all partials of the sequence converge uniformly to the corresponding partial of the limit function. *Note:* We do not ask that the convergence be simultaneously uniform for all partials. This rather intricate definition of convergence of test functions is designed to produce a very simple calculus of distributions. For example, all distributions will be infinitely differentiable.

Exercise 1.1.1 (Distributions Generalize Functions). Suppose that $f(x)$ is a **locally integrable function**; i.e., f is Lebesgue integrable over every bounded Lebesgue measurable subset of $\mathbb{R}^m$. Then f defines a distribution T_f via

$$T_f(g) = \int f(x)g(x)dx \quad \text{for any } g \text{ in } \mathcal{D}.$$

Here $dx =$ the usual Lebesgue measure on $\mathbb{R}^m$. The exercise is to prove that T_f is indeed a distribution. You should also show that two locally integrable functions f_1 and f_2 define the same distribution if and only if $f_1 = f_2$ almost everywhere.

Exercise 1.1.2 (The Dirac Distribution). Define $\delta(g) = g(0)$ for each test function g in $\mathcal{D}$. Show that delta is indeed a distribution (i.e., continuous and linear).

The second definition of distribution presents T as an equivalence class of Cauchy sequences of locally integrable functions. Here again convergence is defined with respect to test functions. That is, f_n converges to f means that $\int f_n g \, dx$ converges to $\int fg \, dx$ for every test function g. You can find more details in Korevaar [371], for example.

The Dirac delta distribution corresponds to the following type of sequence, for example.

Definition 1.1.1. A **Dirac sequence of positive type** $K_n : \mathbb{R}^m \to \mathbb{C}$ has the following properties:

(1) K_n is integrable over $\mathbb{R}^m$.
(2) K_n is nonnegative.
(3) $\int K_n(x)\,dx = 1$.
(4) For every $\varepsilon, \delta > 0$, there exists an $N \in \mathbb{Z}^+$ such that when $n \geq N$:

$$\int\limits_{\|x\| \geq \delta} K_n(x)dx < \varepsilon.$$

It is easy to show (see Lang [386, pp. 211–213]) that if f is bounded and continuous on $\mathbb{R}^m$, then the convolution $(K_n * f)(x) = \int K_n(x-y)f(y)dy$ approaches $f(x)$ uniformly, as n goes to infinity, for x in any compact set.

Examples of Dirac Sequences for $\mathbb{R}^m = \mathbb{R}$

Example 1.1.1 (The Fejér Kernel).

$$K_n(x) = \begin{cases} \frac{1}{n}\left(\frac{\sin(n\pi x)}{\sin(\pi x)}\right)^2, & \text{for } |x| \leq \frac{1}{2}, \\ 0, & \text{otherwise.} \end{cases}$$

This kernel can be used to show that the Fourier series of a continuous periodic function can be Cesàro summed to converge uniformly to the function (see Sect. 1.3 and Lang [386, p. 233]).

Example 1.1.2 (The Landau Kernel).

$$c_n = \left(\int_{-1}^{+1}(1-x^2)^n dx\right)^{-1}.$$

$$L_n(x) = \begin{cases} c_n\left(1-x^2\right)^n, & \text{for } |x| \leq 1, \\ 0, & \text{otherwise.} \end{cases}$$

This kernel can be used to show that any continuous function on a finite interval can be uniformly approximated by polynomials since $L_n * f$ is a polynomial (see Lang [386, p. 214]).

Example 1.1.3 (The Gauss Kernel). For $t > 0$,

$$G_t(x) = \frac{1}{t\sqrt{2\pi}}\exp[-x^2/(2t^2)], \quad \text{as } t \text{ approaches zero from the right.}$$

This kernel is the fundamental solution of the heat equation in one space variable (*see* Sect. 1.2). It is also the density for the normal distribution of mean 0 and standard deviation $\sqrt{t}$. See Fig. 1.2 for some graphs of G_t, for t between 0.1 and 1.

Exercise 1.1.3. (a) Check that the three preceding examples are Dirac sequences of positive type. Show that the convolution $K_n * f$ of any of these kernels with a continuous function f approaches f uniformly on compact sets as n goes to infinity. For Example 1.1.3, the index n is replaced by the positive real number t which approaches 0 from the right.

(b) Then show that any continuous function of period 1 can be uniformly approximated by polynomials $P(x)$ as well as by trigonometric polynomials $P(\exp(2\pi iax))$.

To obtain Dirac sequences for $\mathbb{R}^m$, $m > 1$, take $K_n(x) = K_n(x_1)\cdots K_n(x_m)$ for $x = (x_1, \ldots x_m)$.

An example of a sequence approaching delta distributionally which is not a Dirac sequence of positive type is the Dirichlet kernel, used to show that the Fourier series of a continuously differentiable function converges to the function (see Exercise 1.3.1 of Sect. 1.3 and Lang [386, p. 237]).

It is clear that T_f in Exercise 1.1.1 corresponds to a sequence of functions. Take every element of the sequence to be f itself.

It is easy to define the restriction, sum, scalar multiple, translate, support, and derivative of a distribution. For example, the **support of a function** f is the smallest closed set outside of which f vanishes. We say that a distribution T is **zero on an open set** U in $\mathbb{R}^m$ if $T(g) = 0$ for all test functions g with support in U. Then the **support of a distribution** is the smallest closed set outside of which the distribution is zero.

Exercise 1.1.4. Show that the support of the Dirac distribution is $\{0\}$.

Integration by parts says that if f' is locally integrable and g is a test function of one variable, then

$$T_{f'}(g) = \int f'g = -\int fg' = -T_f(g').$$

This suggests that the **partial derivative** $\partial T/\partial x_i$ of a distribution T should be defined by setting

$$(\partial T/\partial x_i)\,(g) = -T(\partial g/\partial x_i)$$

for any test function g. One must check that $-T(\partial g/\partial x_i)$ has a continuous linear dependence on the test function g. This follows from the way that convergence of test functions was defined. Thus all distributions are infinitely differentiable and all mixed partials of distributions must be equal.

Exercise 1.1.5 (The Distributional Derivative of the Heaviside Step Function Is the Dirac Delta Distribution on $\mathbb{R}^m = \mathbb{R}$). Define the **Heaviside function** by

$$H(x) = \begin{cases} 0, & x < 0, \\ 1, & x > 0. \end{cases}$$

Show that $H' = \delta$, as distributions.

The following exercise is useful in the study of PDEs involving the Laplace operator Δ. Suppose that E is a fundamental solution of Laplace's equation (also known as the free space Green's function for the Laplacian); i.e. $\Delta E = \delta$ (a distributional differential equation). Then, if we are given a sufficiently nice function f and we seek to solve the PDE $\Delta u = f$ for the unknown function u, it will easily be seen that $u = f * E$ (cf. Theorem 1.1.1, which follows). It is harder to find the Green's functions satisfying various sorts of boundary conditions (see Garabedian [197]).

Exercise 1.1.6 (The Fundamental Solution of the Laplacian in $\mathbb{R}^m$, $m \geq 3$). Show that in the sense of distributions: if $\Delta = \partial^2/\partial x_1^2 + \cdots + \partial^2/\partial x_m^2$, then

$$\Delta(\|x\|^{2-m}) = (2-m)S_m\delta, \text{ where } S_m = \text{ area of unit sphere in } \mathbb{R}^m.$$

Thus the fundamental solution for the Laplacian is $E_m(x) = c_m\|x\|^{2-m}$,

$$c_m = [S_m(2-m)]^{-1}.$$

Hint. *Let T be the distribution of the function E_m in the sense of Exercise 1.1.1. You must show that $(\Delta T)(g) = T(\Delta g) = g(0)$ for each test function g. Now $T(\Delta g) = c_m \int \|x\|^{2-m}\Delta g\, dx$. To evaluate this integral, apply Green's theorem to the region obtained by removing a ball of radius r from $\mathbb{R}^m$. Then let r approach zero. In Sect. 1.2, we will learn another way to do this.*

Exercise 1.1.7 (The Area of the Unit Sphere in $\mathbb{R}^m$). Show that

$$\text{area}\{x \in \mathbb{R}^m \mid \|x\| = 1\} = 2\Gamma\left(\frac{1}{2}\right)^m / \Gamma(m/2).$$

Hint. *Consider the integral of* $\exp\left(-\|x\|^2\right)$ *in polar coordinates and recall Euler's formula for the gamma function (see Lebedev [401]).*

It is useful to note **a few general facts about distributions** that we do not have the space to prove. There are three such results.

Three Facts

(1) Any distribution supported by $\{0\}$ is a finite linear combination of partial derivatives of delta.

(2) Any distribution with compact support extends to a continuous linear functional on the space $C^\infty(\mathbb{R}^m)$ of all infinitely differentiable functions on $\mathbb{R}^m$. Here convergence of a sequence of C^∞ functions means uniform convergence of partials on compacta.

(3) Every distribution is locally the result of applying a distributional differential operator (with constant coefficients) to a continuous function. Here "locally" means: when restricted to test functions with support in a given compact set.

Remarks On Proofs.

(1) See Lang [387, p. 448].
(2) See Friedlander [186, p. 35].
(3) See Gelfand and Shilov [206, Chap. 2]. The proof can be sketched as follows. Now test functions with support in a compact set S are the intersection over p of functions which are continuously differentiable of order up to p on S. This dualizes to say that a p must exist so that T will be a continuous linear functional on functions continuously differentiable of order up to p. Then one uses the Hahn–Banach theorem, the Riesz representation theorem, and integration by parts. We shall use this result in the discussion of convolution and Fourier transform of distributions.

One defines **convergence** of a sequence T_n of distributions to a distribution T as n approaches infinity to mean $T_n(g)$ approaches $T(g)$ for every test function g. When thinking of functions as distributions, as in Exercise 1.1.1, this type of convergence is often called **weak**. The Lebesgue dominated convergence theorem gives an easy condition for weak convergence of sequences of functions. It is of course much easier for a sequence of functions to converge in the sense of distributional convergence than to converge pointwise. Note that if K_n is a Dirac sequence, then T_{K_n} approaches δ as $n \to \infty$.

It is easy to see that the operations of differentiation and passing to the limit can always be interchanged distributionally. This is false, of course, for sequences of functions. Just as easily, one can see that very weak conditions guarantee the convergence of series of distributions (for example, see Schwartz [565, p. 97]). In Theorem 1.3.3 of Sect. 1.3, we shall see some examples of Fourier series of distributions.

In general, it is impossible to define the **product of two distributions** without abandoning associativity, for we shall see that on $\mathbb{R}$ we have $x\delta = 0$. But then

$$\left(\frac{1}{x}x\right)\delta = \delta \quad \text{and} \quad \frac{1}{x}(x\delta) = 0.$$

Note also that if f is locally integrable, there is no reason for f^2 to be locally integrable; e.g., $f(x) = |x|^{-1/2}$. The nastier T is, the nicer S must be for TS to make sense. We shall confine ourselves to the following situation.

Definition 1.1.2. If T is a distribution and α is any infinitely differentiable function on $\mathbb{R}^m$, define the distribution αT by $\langle \alpha T, g\rangle = \langle T, \alpha g\rangle$, using the notation $T(g) = \langle T, g\rangle$, for any test function g.

Exercise 1.1.8 (Products of Infinitely Differentiable Functions α and Distributions T).

(a) Show that

$$\frac{\partial}{\partial x_i}(\alpha T) = \frac{\partial \alpha}{\partial x_i} T + \alpha \frac{\partial T}{\partial x_i}.$$

(b) Show that $\alpha\delta = \alpha(0)\delta$.

One must similarly make some restrictions in order to define the convolution of two distributions. The idea of convolution is quite important. For example, solutions to partial (and ordinary) differential equations are often convolutions (cf. Exercise 1.1.6 and the next section). In probability theory, the density function of the sum of two independent random variables is the convolution of the two density functions (see, Exercise 1.2.22 in Sect. 1.2 or Apostol [10, Vol. II, p. 552]).

In order to figure out what the convolution of two distributions should be, we must recall the definition of **convolution** $f{*}g$ **of two functions** $f, g : \mathbb{R}^m \to \mathbb{C}$, assuming both of the functions are locally integrable and that at least one has bounded support:

$$(f*g)(x) = \int_{\mathbb{R}^m} f(x-y)g(y)dy = \int_{\mathbb{R}^m} f(y)g(x-y)dy. \tag{1.2}$$

The convolution of f, g in $L^1(\mathbb{R}^m) =$ the Lebesgue integrable functions also makes sense.

One can define the **convolution of two distributions** as the equivalence class of Cauchy sequences obtained by convolving the functions representing the two distributions, assuming that one of the distributions has bounded support. The definition as a continuous linear functional goes as follows.

Definition 1.1.3. Suppose that S and T are distributions and that S has bounded support. Then define the distribution $T*S$ as follows using the notation $T(g) = \langle T, g\rangle$:

$$\langle T*S, g\rangle = \langle T(y), \langle S(x), g(x+y)\rangle\rangle \quad \text{for each test function } g.$$

Here we are abusing notation by writing distributions as if they were functions in order to keep track of variables.

In order to see that the preceding definition of convolution of distributions makes sense, one must show that $\langle S(x), g(x+y)\rangle$ is a test function as a function of y. To see this, use the fact that S is the distributional derivative of a continuous function (locally for test functions vanishing off a given compact set).

Exercise 1.1.9. (a) Check that convolution of functions is associative and commutative, assuming that all the functions are in $L^1(\mathbb{R}^m)$. Show that convolution of a Lebesgue integrable function with a differentiable function produces a differentiable function.

(b) Suppose that f and g are locally integrable functions and that g has compact support. Show that $T_{f*g} = (T_f)*(T_g)$, using the notation of Exercise 1.1.1.

Theorem 1.1.1 (Properties of Convolution).

(a) ***Regularization.*** *The convolution of a distribution of bounded support and a test function is a test function.*

(b) $T * \delta = T$, $T*(\frac{\partial \delta}{\partial x_i}) = \frac{\partial T}{\partial x_i}$, *with* $\frac{\partial T}{\partial x_i}$ *denoting the partial derivative of* T *with respect to* x_i.

(c) $T*S = S*T$ *if the support of* S *or of* T *is bounded.*

(d) $(T*S)*R = T*(S*R)$ *if the supports of any two of these distributions are bounded.*

(e) $\frac{\partial}{\partial x_i}(T*S) = (\frac{\partial T}{\partial x_i})*S = T*(\frac{\partial S}{\partial x_i})$ *if the support of* S *or* T *is bounded.*

(f) Suppose that T_n *approaches* T *in the sense of distributions. Then* T_n*S *approaches* $T*S$ *as* n *goes to infinity, provided that one of the following conditions holds:*

(i) the T_n *have uniformly bounded supports,*

(ii) S *has bounded support.*

(g) Define the ***translate***[1] *of a distribution* T *by a vector* $a \in \mathbb{R}^m$ *as*

$$T_a(g) = T(g_{-a}) \quad \textit{where} \quad g_a(x) \ = g(x+a).$$

Then $T*\delta_a = T_a$ *and* $\delta_a*\delta_b = \delta_{a+b}$.

Exercise 1.1.10. Prove Theorem 1.1.1.

It follows from Theorem 1.1.1 that every distribution is the limit of a sequence of test functions. To see this, note that there is a sequence of test functions K_n approaching the Dirac delta distribution. Therefore $T*K_n$ approaches T as n approaches infinity. For we can find K_n with uniformly bounded supports and apply part (f) of Theorem 1.1.1.

Exercise 1.1.11. Show that the support of $(S*T)$ is contained in the closure of the set of points $x+y$, with x in the support of S, and y in the support of T.

1.2 Fourier Integrals

> Euclidean methodology has developed a certain obligatory style of presentation. I shall refer to this as "deductivist style." This style starts with a painstakingly stated list of *axioms, lemmas* and/or *definitions*. The axioms and definitions frequently look artificial and mystifyingly complicated. One is never told how these complications arose. The list of axioms and definitions is followed by the carefully worded *theorems*. These are loaded with

---

[1] This notation may be confusing since we are also using subscripts for many other purposes. Reader be alert!

> heavy-going conditions; it seems impossible that anyone should ever have guessed them. The theorem is followed by the *proof*. . . .
>
> Deductivist style hides the struggle, hides the adventure. The whole story vanishes, the successive tentative formulations of the theorem in the course of the proof-procedure are doomed to oblivion while the end result is exalted into sacred infallibility.
>
> —From Lakatos [384, p. 142].

1.2.1 Fourier Transforms

We want to understand the Fourier transform of a distribution on $\mathbb{R}^m$, but it makes sense to start off with Fourier transforms of Schwartz functions. Our study of Fourier transforms on Euclidean space will be sketchy. For more details, the reader could consult any of the following references: Benedetto [34], Bochner [43], Yvonne Choquet-Bruhat et al. [87], Dieudonné [134], Dym and McKean [147], Körner [372], Lang [386, 387], Marks [453], Maurin [458], Schwartz [565, 566], Stade [617], Stein and Shakarchi [635], Stein and Weiss [636], Strichartz [643], Titchmarsh [678], Zygmund [757].

Definition 1.2.1. The **Schwartz space** $\mathcal{S}$ is the space of all infinitely differentiable functions $f:\mathbb{R}^m \to \mathbb{C}$ such that $\left|x^a D^b f\right|$ is bounded for all $a,b \in \mathbb{Z}^m$, with $a_j \geq 0$ and $b_j \geq 0$ for all j. Here we use the notation: $a = (a_1,\ldots,a_m)$,

$$\begin{aligned} &x^a = x_1^{a_1} \cdots x_m^{a_m} \quad \text{and} \quad D^b = \frac{\partial^{|b|}}{\partial x_1^{b_1} \cdots \partial x_m^{b_m}}, \\ &\text{with} \quad |b| = b_1 + \cdots + b_m. \end{aligned} \tag{1.3}$$

We will call such functions f **Schwartz functions**.

Definition 1.2.2 (Fourier Transform of Schwartz Functions). Suppose $f \in \mathcal{S}$. Then define the **Fourier transform** $\hat{f}$ by

$$\hat{f}(y) = \int\limits_{\mathbb{R}^m} f(x) \exp\left(-2\pi i\, {}^t x y\right) dx.$$

Here we write $x \in \mathbb{R}^m$ as a column vector and ${}^t x$ as the transpose of that vector. Thus ${}^t xy$ is the inner product of x and y.

Note that the integral defining $\hat{f}(y)$ is easily seen to be convergent.

Theorem 1.2.1 (Properties of the Fourier Transform on the Schwartz Space).

(1) $f \in \mathcal{S}$ *implies that* $\hat{f} \in \mathcal{S}$.

(2) $D^a\left(\hat{f}\right) = \left((-2\pi i x)^a f\right)^{\widehat{}}$, *using the notation (1.3) for differential operators.*

(3) $\left(D^a f\right)^{\widehat{}} = (2\pi i x)^a \hat{f}$.

(4) ***Convolution Theorem:*** $\widehat{(f*g)} = \hat{f}\cdot\hat{g}$, *where convolution is defined by formula (1.2).*

(5) ***Translation:*** *Set* $f_a(x) = f(x+a)$ *for* a, x *in* $\mathbb{R}^m$. *Then*

$$\widehat{f_a}(x) = \exp(2\pi i\,{}^t a x)\hat{f}(x).$$

(6) ***A Function That Is Its Own Fourier Transform:*** *Let* $f(x) = \exp(-\pi\|x\|^2)$. *Then* $\hat{f} = f$.

(7) ***Dilation:*** *Let* u *be a positive real number and set* ${}^u f(x) = f(ux)$ *for* x *in* $\mathbb{R}^m$. *Then*

$$\widehat{({}^u f)}(x) = u^{-m}\hat{f}(u^{-1}x).$$

(8) ***Multiplication Formula:*** $\int \hat{f}g = \int f\hat{g}$.

(9) ***Fourier Inversion:*** *Define* $f^-(x) = f(-x)$. *Then* $\hat{\hat{f}} = f^-$.

All of the functions in the preceding statements are assumed to be Schwartz functions.

Proof. Everything is an exercise except part (9). To prove (9), note that the operation of replacing f by $\hat{\hat{f}}^{-}$ commutes with translation. Thus it suffices to prove (9) at $x = 0$. Then property (8) shows that it suffices to prove (9) for a Dirac sequence of Schwartz functions such as the Gauss kernel from Sect. 1.1. For we have, assuming that $K_n = K_n^- = \widehat{\widehat{K_n}}$,

$$\int \hat{\hat{f}} K_n = \int \hat{f}\,\widehat{K_n} = \int f\widehat{\widehat{K_n}} = \int f\,K_n.$$

The left-hand side of this equality approaches $\hat{\hat{f}}(0)$ and the right-hand side approaches $f(0)$ as n goes to infinity, since K_n approaches the Dirac delta distribution. □

Exercise 1.2.1. Prove property (2) of the Fourier transform by differentiating under the integral sign. Then prove property (3) by integration by parts. Property (1) then follows easily from (2) and (3).

Exercise 1.2.2. Prove property (4) by changing the order of integration.

Exercise 1.2.3. Prove properties (5) and (7) by making the right substitutions.

Exercise 1.2.4 (Fourier Transforms of Gauss Kernels). Show that

$$f(x) = \exp(-\pi\|x\|^2), \quad x \in \mathbb{R}^m,$$

is its own Fourier transform by computing $\widehat{\frac{df}{dx}}(x)$ using integration by parts and then solving the resulting differential equation. Deduce property (6) of Theorem 1.2.1.

Then let G_t be the Gauss kernel (Example 1.1.3 of Sect. 1.1). Show that $G_t = G_t^- = \widehat{\widehat{G_t}}$ using the seventh property of the Fourier transform.

Exercise 1.2.5 (More Properties of the Fourier Transform). Show that if f and g are Schwartz functions:

(a) $f \mapsto \hat{f}$ is one-to-one, linear from $\mathcal{S}$ onto $\mathcal{S}$.
(b) $\widehat{(fg)} = \hat{f} * \hat{g}$.
(c) Define an inner product for f,g in $\mathcal{S}$ by $(f,g) = \int f\,\overline{g}$. Then $(f,g) = (\hat{f},\hat{g})$ **(Parseval's identity).** And, setting $\|f\|_2 = (f,f)^{1/2}$, we have the **Plancherel identity** $\|f\|_2 = \|\hat{f}\|_2$.

Mathematicians call the last statement of Exercise 1.2.5 the **Plancherel theorem**. And we will see many generalizations of it in the succeeding chapters. Plancherel proved it in 1910. However, Rayleigh used this result in his study of blackbody radiation in 1889 (see Rayleigh [537]). Thus it might be more accurately called Rayleigh's theorem. In later chapters this result will be equivalent to the determination of the spectral or Plancherel measure in the inversion formulas for the generalized Fourier transforms that we shall study.

Now we should turn to Fourier integrals of nastier functions than Schwartz functions. The space of all Schwartz functions is dense in the space $L^2(\mathbb{R}^m)$ of all Lebesgue square integrable functions on $\mathbb{R}^m$. See Apostol [11], Yvonne Choquet-Bruhat et al. [87], Dieudonné [134, Vol.II], Dym and McKean [134], Kolmogorov and Fomin [366], Korevaar [371], Lang [387], Maurin [458], or any of countless texts that treat the theory of the Lebesgue integral. The book by Dym and McKean is strongly recommended for this purpose.

The continuous linear map of $\mathcal{S}$ onto $\mathcal{S}$ (using the L^2-topology of Exercise 1.2.5) given by $f \mapsto \hat{f}$, has a unique continuous extension to a map on $L^2(\mathbb{R}^m)$. The extended mapping will be one-to-one, onto and a linear Hilbert space isomorphism of $L^2(\mathbb{R}^m)$. It is easy to show that if $f \in L^2(\mathbb{R}^m)$, then $\hat{f}(y)$ is the L^2-limit, as n goes to infinity, of the finite Fourier transforms:

$$\int_{\|x\| \leq n} f(x) \exp(-2\pi i\, {}^t xy)\,dx.$$

Here you need the L^2 dominated convergence theorem. The inverse transform has a similar characterization.

Wiener showed in [736] that it is also possible to give an elegant description of the L^2 Fourier transform using Hermite polynomials. We will discuss this in Sect. 1.3.

If f is in $L^1(\mathbb{R}^m)$; i.e., if f is Lebesgue measurable and $\|f\|_1 = \int |f|$ is finite, then the Fourier transform $\hat{f}(y) = \int f(x) \exp(-2\pi i\, {}^t xy)dx$ exists as a Lebesgue integral. We could not say this for L^2 functions. Thus it appears more natural to discuss Fourier integrals for L^1 functions. However, there is a problem with L^1 functions. The Fourier transform of an L^1 function need not be L^1. A simple example on the real line is given by that in the following exercise.

Exercise 1.2.6 (A Function Where the Fourier Transform Is Not in $L^1(\mathbb{R})$). Define

$$\Pi(x) = \begin{cases} 0, \ |x| > \frac{1}{2}, \\ \frac{1}{2}, \ |x| = \frac{1}{2}, \\ 1, \ |x| < \frac{1}{2}. \end{cases}$$

Show that $\hat{\Pi}(y) = \sin(\pi y)/(\pi y)$. Then show that $\hat{\Pi}$ is not in $L^1(\mathbb{R})$. In fact, this gives us an example of a function that has an improper Riemann integral on $\mathbb{R}$ but is not absolutely integrable and thus is not Lebesgue integrable.

Theorem 1.2.2 (Properties of the Fourier Transform of L^1Functions). *If $f \in L^1(\mathbb{R}^m)$, we have the following facts.*

(1) $\|\hat{f}\|_\infty = \text{l.u.b.}\ |\hat{f}(y)| \leq \|f\|_1$ *(*l.u.b. =*least upper bound).*
(2) $\hat{f} : \mathbb{R}^m \to \mathbb{C}$ *is continuous.*
(3) $\widehat{(f*g)} = \hat{f} \cdot \hat{g}$*.*
(4) ***Riemann–Lebesgue lemma:*** $\lim_{|y|\to\infty} \hat{f}(y) = 0$.
(5) $\hat{f} = 0 \Leftrightarrow f = 0$. *Here we mean* 0 *almost everywhere.*

Discussion. The proofs of these properties are not hard, using Lebesgue dominated convergence, Fubini's theorem, etc. To prove the Riemann–Lebesgue lemma, note that given an L^1 function f, one can find a Schwartz function g that is close to f in L^1 norm. Then $\hat{f}(x)$ and $\hat{g}(x)$ must be close for all x. Now use the fact that $\hat{g}$ approaches zero as x goes to infinity. More details can be found in Dym and McKean [147] or Stein and Weiss [636], for example.

In part (5) of Theorem 1.2.2, we are saying that the Fourier transform defines a 1–1 map of $L^1(\mathbb{R}^m)$ into the continuous functions on $\mathbb{R}^m$ that vanish at infinity. The actual image is a very complicated subalgebra, which has not yet been characterized.

You might well ask how to reclaim a function f in $L^1(\mathbb{R})$ from its Fourier transform. Since $\hat{f}$ need not be in L^1, it is necessary to use some kind of summability procedure. The time-honored method for piecewise smooth functions is outlined in the following exercise. Other methods can be found in Dym and McKean [147, pp. 103–106].

Exercise 1.2.7 (A Dirac Family Associated to Fourier Inversion). Suppose that $f \in L^1(\mathbb{R})$ and f is piecewise continuous (i.e., it can have at most a finite number of jump discontinuities in each finite interval). If the one-sided derivatives of f exist at x in $\mathbb{R}$, show that

$$\begin{aligned} \frac{1}{2}(f(x+) + f(x-)) &= \lim_{r\to\infty} \left(\int_{|y|\leq r} \hat{f}(y) \exp(2\pi i x y) dy \right) \\ &= \lim_{r\to\infty} (f * W_r) \quad \text{with} \quad W_r(x) = \frac{sin(2\pi r x)}{\pi x}. \end{aligned}$$

Here

$$f(x+) = \lim_{w \to x, w > x} f(w) \quad \text{and} \quad f(x-) = \lim_{w \to x, w < x} f(w).$$

Hint. *(cf. Exercise 1.1.1 of Sect. 1.1) Write*

$$\int_{|y| \le r} \hat{f}(y) \exp(2\pi ixy) dy = \int_{|y| \le r} \int_{u=-\infty}^{+\infty} \exp(2\pi iy(x-u)) f(u)\, du\, dy.$$

Then interchange the order of integration.

To complete the solution, you need the Riemann–Lebesgue lemma (Theorem 1.2.2, Part 4) and Exercise 1.2.14 below which says that

$$\int_0^\infty \frac{\sin x}{x} dx = \frac{\pi}{2}.$$

Note. It is easy to extend Exercise 1.2.7 to $\mathbb{R}^m$, since

$$\int_{\substack{y \in \mathbb{R}^m \\ |y_j| \le r_j}} \hat{f}(y) \exp(2\pi i\, {}^t xy) dy = (f * W_r^m)(x),$$

where

$$W_r^m(x) = \prod_{j=1}^m W_{r_j}(x_j) \quad \text{for} \quad x \in \mathbb{R}^m,\ r \in (\mathbb{R}^+)^m.$$

Fourier analysis, however, did not originate in this textbook style of theorems and exercises. At the beginning, there were Fourier's experiments on heat diffusion. And there was much controversy! For example, Lagrange prevented the publication of Fourier's 1807 paper on the subject. See Grattan-Guinness and Ravetz [230] for the first published version of Fourier's paper, as well as an interesting discussion of Fourier's life and work. There was tremendous rivalry between mathematicians such as Fourier, Cauchy, and Poisson. Colossal arguments took place over the truth of the proposition that "any" function can be represented by its Fourier series. Mathematicians made insulting reviews of each other's papers. Out of real physical questions and equally real mathematical difficulties came modern analysis—the Riemann and Lebesgue integrals and, later, distributions. See also Riemann [542, pp. 227–271], for the fascinating history of Fourier analysis. Other historical references are Burkhardt [72] and Hilb and Riesz [296]. Another interesting reference is I. Lakatos [384, pp. 128–131].

Example 1.2.1. ***[Heat Flow on an Infinite Rod]*** The problem of heat flow on an infinite rod can be posed mathematically as follows, making many simplifying assumptions about the rod:

Find $u(x,t)$, $x \in \mathbb{R}$, $t > 0$, satisfying the heat equation:

$$\begin{cases} \frac{\partial u}{\partial t} = a^2 \frac{\partial^2 u}{\partial x^2}, & \text{for } x \in \mathbb{R}, \quad t > 0, \\ u(x,0) = f(x), & \text{where } f(x) = \text{given initial heat distribution.} \end{cases}$$

For a derivation see Dym and McKean [147, p. 61], or Vladimirov [706, p. 31], for example. Or you may want to read Fourier's original paper with its various approximations to the "correct" partial differential equation in Grattan-Guinness and Ravetz [230]. The constant a in the differential equation depends on the specific heat of the (uniform) material composing the rod.

To solve the problem, formally Fourier transform both sides of the partial differential equation with respect to x. Use the fact that the Fourier transform takes differentiation to multiplication by $2\pi ix$ (property 3 of Theorem 1.2.1). You obtain

$$\frac{\partial \hat{u}}{\partial t} = -4\pi^2 x^2 a^2 \hat{u}.$$

Therefore, $\hat{u} = \hat{f} \exp(-4\pi^2 a^2 x^2 t)$.

Using Exercise 1.2.4, we see that the Fourier transform of $\exp[-(2\pi a x)^2 t]$ as a function of x is

$$G_v(x) = (2a\sqrt{\pi t})^{-1} \exp[-x^2/(4a^2 t)], \quad \text{where } G_v = \text{Gauss kernel}, \quad v = a(2t)^{1/2}.$$

Now use the convolution property of the Fourier transform (Theorem 1.2.2, Part 3) and Fourier inversion to see that

$$u(x,t) = \hat{\hat{u}}(-x,t) = G_v * f = \left(G_{2a^2t} * f\right)(x).$$

Here convolution is over the variable x.

The entire procedure of the preceding example must now be justified, since we did not know at the beginning that it was legal to use Fourier transform properties of $u(x,t)$, as it was an unknown function. Rather weak assumptions on f will suffice to make $G_{a(2t)^{1/2}}(x) * f$ a solution of our problem, since the Gauss kernel is infinitely differentiable and a Schwartz function. Moreover, G_t is a Dirac sequence of positive type as t approaches 0 from the right. We are using Exercises 1.1.3 and 1.1.10 from Sect. 1.1.

We shall see at the end of this section that the Gauss kernel is also of central importance in probability and statistics.

Exercise 1.2.8 (D'Alembert's Solution of the Wave Equation). Consider the displacement $u(x,t)$ of an infinite homogeneous vibrating string at position x on the real axis and time t. Making enough simplifying assumptions, one finds that $u(x,t)$ must satisfy the wave equation (see, e.g., Dym and McKean [147, p. 62]):

$$\begin{cases} \frac{\partial^2 u}{\partial t^2} = a^2 \frac{\partial^2 u}{\partial x^2}, & \text{for } x \in \mathbb{R}, \quad t > 0, \\ u(x,0) = f(x), & f(x) = \text{given initial heat distribution,} \\ \frac{\partial u}{\partial t}(x,0) = g(x), & g(x) = \text{given initial velocity.} \end{cases}$$

Apply similar methods to those that we used in the heat equation example to obtain d'Alembert's solution:

$$u(x,t)=\frac{1}{2}[f(x+at)+f(x-at)]+\frac{1}{2a}\int_{x-at}^{x+at}g(u)du.$$

Now that we have briefly reviewed the theory of the Fourier transform for rather nasty functions, it is time to describe the theory of the Fourier transform for distributions. The multiplication formula (property 8 of Theorem 1.2.1) suggests that the Fourier transform $\hat{T}$ of a distribution T should be defined by the equation $\hat{T}(g)=T(\hat{g})$ for all test functions g. The only problem is that when g and $\hat{g}$ are both test functions, then g must be identically zero, as we shall soon see. It is not, in fact, surprising that the support of $\hat{g}$ should be like that of $\exp(2\pi ix)$.

The moral is that we cannot easily define the Fourier transform of every distribution T. Instead we must restrict ourselves to tempered distributions. There is another way out of this dilemma which makes $\hat{T}$ a new sort of generalized function (see Gelfand and Shilov [206]).

Definition 1.2.3. A distribution T is said to be a **tempered distribution** if it extends to a continuous linear functional on the space $\mathcal{S}$ of Schwartz functions.

In order to understand what we mean by continuity in the definition above, we must define what we mean by **convergence of a sequence of Schwartz functions** g_n to a Schwartz function g. This means that all of the sequences $x^aD^b(g_n-g)$ converge to zero uniformly on $\mathbb{R}^m$ for all a,b in $(\mathbb{Z}^+)^m$, using notation (1.3) which was set up at the beginning of this section.

Exercise 1.2.9. Show that the Fourier transform defines a continuous linear map from $\mathcal{S}$ onto $\mathcal{S}$, using the preceding definition of continuous.

Definition 1.2.4. Suppose that T is a tempered distribution. Define the **Fourier transform** $\hat{T}$ by $\hat{T}(g)=T(\hat{g})$, for all Schwartz functions g.

Exercise 1.2.10 (Distributions Tempered and Not So Tempered).

(a) Show that the following distributions are all tempered:
the distribution of a bounded or Lebesgue integrable function,
any distribution with bounded support.

(b) Show that e^x does not define a tempered distribution on $\mathbb{R}$.

Theorem 1.2.3 (Properties of the Fourier Transform of Tempered Distributions).

(1) The Fourier transform of a tempered distribution T with bounded support is a function:

$$V(s)=T(exp(-2\pi i\,{}^tsx)).$$

Moreover, V may be continued to all complex vectors $s \in \mathbb{C}^m$ as an entire function with bound

$$|V(s)| \leq Ce^{A\|\mathrm{Im}\, s\|}\left(1+\|s\|^N\right),$$

for some A and N. The converse of this theorem is also true and is called the Schwartz–Paley–Wiener theorem.

(2) ***Fourier Transform Exchanges Differentiation and Multiplication by Polynomials.*** *Using the notation introduced in the definition of the Schwartz space, we have:*

$$D^a\left(\hat{T}\right) = ((-2\pi i x)^a T)\widehat{\ } \quad \text{and} \quad \widehat{(D^a T)} = (2\pi i x)^a \hat{T}.$$

(3) ***The Convolution Theorem:*** *Suppose that S and T are two distributions with bounded supports. Their Fourier transforms are functions and*

$$\widehat{(S * T)} = \hat{S}\hat{T}.$$

(4) ***Interchange of Unit for Convolution and Unit for Multiplication:***

$$\hat{\delta} = 1.$$

(5) ***Fourier Inversion:*** *Suppose that T is a tempered distribution and g is a Schwartz function. Define T^- by $T^-(g) = T(g^-)$, where $g^-(x) = g(-x)$ for Schwartz functions g. Then*

$$T = (\hat{\hat{T}})^-.$$

A Few Proofs.

(1) Note that $V(s)$ makes sense because T has compact support and thus extends to the space of infinitely differentiable functions. For the same reason, there is a constant coefficient differential operator D with $T = DF$ for some continuous function F. Then D is a sum of terms involving differential operators L of even or odd order. This allows us to move L around to g with a ±. Thus, for Schwartz functions g, we have (setting $T(g) = \langle T, g\rangle$):

$$\begin{aligned}\langle \hat{T}, g\rangle = \langle T, \hat{g}\rangle &= \sum \langle LF, \hat{g}\rangle \\ &= \sum \pm \int F(x)\left(L_x \int \exp(-2\pi i\, {}^t s x) g(s) ds\right) dx \\ &= \langle\langle T(x), \exp(-2\pi i\, {}^t s x)\rangle, g(s)\rangle,\end{aligned}$$

upon interchanging orders of integrals. Proofs of the Paley–Wiener theorem can be found in Dym and McKean [147] and Schwartz [566, Chap. VI].

Table 1.1 A short table of Fourier transforms of one variable

$f(x),\ x \in \mathbb{R}$	$\hat{f}(y)$
$e^{-\pi x^2}$	$e^{-\pi y^2}$
$\frac{1}{2}e^{-\lvert x\rvert}$	$\frac{1}{1+4\pi^2y^2}$
$\Pi(x) = \begin{cases} 0, & \lvert x\rvert > 1/2 \\ 1/2, & x = 1/2 \\ 1, & \lvert x\rvert < 1/2 \end{cases}$	$\frac{\sin(\pi y)}{\pi y}$
$\delta(x)$	1
$\frac{1}{2}\left(\delta(x-\frac{1}{2}) + \delta(x+\frac{1}{2})\right)$	$\cos(\pi y)$
$\left(q^2+x^2\right)^{-s},\ q \neq 0,\ \operatorname{Re} s > 0$	$\begin{cases} \frac{2\sqrt{\pi}}{\Gamma(s)}\left\lvert\frac{\pi y}{q}\right\rvert^{s-\frac{1}{2}} K_{s-\frac{1}{2}}(2\pi\lvert yq\rvert), & y \neq 0 \\ \frac{\Gamma(\frac{1}{2})\Gamma(s-\frac{1}{2})}{\Gamma(s)} q^{1-s}, & y = 0 \end{cases}$

(3)

$$\begin{aligned} \langle \widehat{S*T}, g\rangle &= \langle S*T, \hat{g}\rangle = \langle S(y), \langle T(x), \hat{g}(x+y)\rangle\rangle \\ &= \left\langle S(y), \left\langle T(x), \widehat{\left(\exp(-2\pi i\,{}^t yx)g(x)\right)}\right\rangle\right\rangle \\ &= \left\langle \hat{T}(x), g(x)\left\langle S(y), \exp(-2\pi i\,{}^t yx)\right\rangle\right\rangle = \left\langle \hat{S}(x)\hat{T}(x), g(x)\right\rangle. \qquad \square \end{aligned}$$

Exercise 1.2.11. Prove parts (2), (4), (5) of Theorem 1.2.3.

It is clear from part (1) of Theorem 1.2.3 that if g and $\hat{g}$ are both test functions, then g must be zero everywhere, since a complex analytic function of one variable which is supported on a bounded infinite set must be identically zero (see Ahlfors [3, p. 127]).

Exercise 1.2.12. Verify Table 1.1.

Hint. *For the last formula, you need some properties of the gamma and K-Bessel functions (see Lebedev [401] or see the first exercise in the second section of Chap. 3).*

In the (good?) old days one needed a long table of Fourier transforms such as Erdélyi et al. [165] or Oberhettinger [500, 501]. These days many computer packages such as Mathematica have such tables built in.

1.2.2 Applications to Partial Differential Equations

Next let us consider some applications of Fourier transforms of distributions to partial differential equations. The general idea is the same as that which we used to solve the heat equation earlier. Suppose that D is a constant coefficient partial differential operator and S is a tempered distribution. To solve the equation

$$DT = S$$

for T, try Fourier transforming the equation. You get

$$\widehat{DT} = M\hat{T} = \hat{S},$$

where M is multiplication by a polynomial. So you want to write

$$\hat{T} = \hat{S}/M.$$

Then Fourier inversion will find T. This makes perfect sense if the polynomial M never vanishes. But otherwise there is much work to do in order to legalize division by M. This has been done by Hörmander [303] and Łojasiewicz [426]. See F. G. Friedlander [186, pp. 139ff], for a discussion of the theorem of Malgrange and Ehrenpreis to the effect that $DE = \delta$ can be solved for E.

In the examples that follow we will Fourier transform the *PDE* with respect to some subset of the variables. This will reduce the problem to an ODE. For more examples of this type, see Vladimirov [706].

Exercise 1.2.13 (Fundamental Solution of the Heat Equation). We want to solve

$$\frac{\partial E}{\partial t} - a^2 \Delta E = \delta(x,t), \qquad x \in \mathbb{R}^m,\ t > 0, \qquad \Delta = \frac{\partial^2}{\partial x_1^2} + \cdots + \frac{\partial^2}{\partial_{X_m^2}}.$$

Claim. $E(x,t) = (4a^2\pi t)^{-m/2} H(t) \exp[-\|x\|^2/(4a^2 t)]$, where H is Heaviside's step function

$$H(t) = \begin{cases} 1, \ t \geq 0, \\ 0 \ \ t < 0. \end{cases}$$

Hint. *Imitate Example 1.2.1. Use* $\hat{\hat{\delta}} = 1$ *and the fact that the distributional ordinary differential equation*

$$S' + cS = \delta(t) = H'(t)$$

has the solution $S(t) = H(t)\exp(-ct)$. *Prove this too.*

Remarks. Note that $E(x,t)$ in Exercise 1.2.13 is positive for all positive t at every point x in space. Thus heat must be diffused with infinite velocity! This means that the heat equation does not appear to be a realistic model of heat transfer. The transport equation appears to be better, but we shall not discuss it here (see Vladimirov [706]).

Example 1.2.2 (Fundamental Solution of Laplace's Equation in $\mathbb{R}^3$*).* We want to solve the following PDE in $\mathbb{R}^3$:

$$\Delta E = \delta.$$

Claim. $E(x) = -(4\pi\|x\|)^{-1}$.

Discussion. Formally Fourier transform the *PDE* with respect to x. One obtains: $-4\pi^2\|x\|^2\hat{E} = 1$. It is possible to justify thinking of $\|x\|^{-2}$ as a tempered distribution since the behavior at 0 and ∞ is not really so bad. Thus $E(-x)$ is the Fourier transform of $(-4\pi^2\|x\|^2)^{-1}$. Suppose that g is a Schwartz function and write, as usual, $T(g) = \langle T, g\rangle$. Then

$$\begin{aligned}\langle E(-x), g\rangle &= \langle(-4\pi^2\|x\|^2)^{-1}, \hat{g}\rangle \\ &= \lim_{R\to\infty}\int_{|x|\le R}(-4\pi^2\|x\|^2)^{-1}\int g(u)\exp(-2\pi i\,{}^t xu)du\,dx.\end{aligned}$$

Interchange the order of integrals and use polar coordinates on the integral over the x variable. Then the x integral becomes:

$$\begin{aligned}&-\frac{2\pi}{4\pi^2}\int_{r=0}^{R}\int_{\theta=0}^{\pi}\exp(-2\pi i\|u\|r\cos\theta)\sin\theta\,d\theta\,dr \\ &\qquad = -\frac{1}{2\pi^2\|u\|}\int_0^{R'}\frac{\sin r}{r}\,dr.\ (R' = 2\pi\|u\|R)\end{aligned}$$

By Exercise 1.2.14, this completes the proof that $\langle E(x), g\rangle = \langle -(4\pi\|x\|)^{-1}, g\rangle$.

Exercise 1.2.14 (An Improper Riemann Integral That Is Not Absolutely Convergent). Show that

$$\int_0^{\infty}\frac{\sin r}{r}dr = \frac{\pi}{2}.$$

Explain why the integral does not converge absolutely.

Exercise 1.2.15 (Fundamental Solution of the Wave Equation in $\mathbb{R}^3$). Define the simple surface layer distribution for a sphere

$$S_R = \left\{x\in\mathbb{R}^3 \,\middle|\, \|x\| = R\right\},$$

by

$$\delta_{S_R}(g) = \int_{x\in S_R} g(x)dA, \quad dA = \text{ surface area element on } S_R.$$

Then show that the solution to the distributional equation $\Box_a T = \delta(x,t)$, where $x\in\mathbb{R}^3$, $t\in\mathbb{R}$, and

$$\Box_a = \left(\frac{\partial^2}{\partial t^2} - a^2 \left(\frac{\partial^2}{\partial x_1^2} + \frac{\partial^2}{\partial x_2^2} + \frac{\partial^2}{\partial x_3^3} \right) \right),$$

is $T = (4\pi a^2 t)^{-1} H(t) \delta_{S_{at}}(x)$.

Remarks. The last exercise can be used to derive Kirchoff's classical formula for the solution of the **Cauchy problem**:

$$\begin{cases} \Box_a u = h(x,t), \quad x \in \mathbb{R}^3, t > 0, \\ u(x,0) = f(x),\ u_t(x,0) = g(x). \end{cases}$$

assuming that f, g, h are all sufficiently smooth (see Vladimirov [706]).

Huyghens' principle also follows from the last exercise, since a disturbance $\delta(x,t)$ propagates as a spherical wave along the spherical surface $\|x\| = at$, moving with speed a. After the wave passes a given point, there is no more disturbance. This principle fails in $\mathbb{R}$ and in even dimensional Euclidean space (see Exercise 1.2.8 and Vladimirov [706] or Dym and McKean [147]). Clearly, Flatland would be a very noisy place!

It is possible to give many interpretations of the Fourier inversion formula that will aid us in our search for non-Euclidean analogues. The most important interpretation for this book is that Fourier inversion provides a spectral resolution of the Laplace operator on $\mathbb{R}^m$. That is, we are investigating **the spectral theorem for a very special unbounded self-adjoint operator** Δ on the Hilbert space $L^2(\mathbb{R}^m)$. For the Fourier inversion formula says that a nice function $f : \mathbb{R}^m \to \mathbb{C}$ has a representation as an integral of elementary eigenfunctions

$$e_a(x) = \exp(2\pi i\, {}^t a x),\ a, x \in \mathbb{R}^m, \tag{1.4}$$

of the Euclidean Laplacian:

$$\Delta = \frac{\partial^2}{\partial x_1^2} + \cdots + \frac{\partial^2}{\partial x_m^2}.$$

For if $f : \mathbb{R}^m \to \mathbb{C}$ is Schwartz, we have the spectral resolution

$$f(x) = \int_{\mathbb{R}^m} \hat{f}(y) e_y(x) dy \quad \text{with} \quad \hat{f}(y) = (f, e_y) = \int_{\mathbb{R}^m} f(u) \overline{e_y(u)} du. \tag{1.5}$$

Note that the eigenfunctions e_a are not in the L^2-space involved. This makes the elaboration of the spectral theorem rather intricate (see Maurin [458]). It is our aim to obtain analogues of Fourier inversion for symmetric spaces such as the non-Euclidean upper half-plane of Chap. 3. We give more information on the spectral theorem in Sect. 2.2. For a discussion of the general spectral theorem the reader could also look at Maurin [458], Reed and Simon [540], or Yosida [748].

Another possible interpretation of the Fourier inversion formula comes from the theory of group representations. In this theory one views the eigenfunctions $e_a(x)$ in the preceding paragraph as irreducible unitary representations of the additive group $\mathbb{R}^m$, since $e_a(x+y) = e_a(x)e_a(y)$. We will not emphasize this aspect here, although we give an introduction to group representations in Sect. 2.2. If the reader is interested in group representations there are many good references, such as Dym and McKean [147], Hamermesh [246], Knapp [356], Mackey [442–444], Talman [655], Michelle Vergne [697], and Vilenkin [704].

Exercise 1.2.16 (Heisenberg's Uncertainty Principle). Suppose that $\|f\|_2 = 1$. Then since $\|f\|_2 = \|\hat{f}\|_2$, both $|f(t)|^2$ and $\left|\hat{f}(w)\right|^2$ can be viewed as densities of probability distributions. Translation of f just leads to a phase shift in $\hat{f}$ (and similarly for translation of $\hat{f}$). So we can assume that the means vanish; i.e.,

$$\int t|f(t)|^2 dt = 0 \quad \textit{and} \quad \int w\left|\hat{f}(w)\right|^2 dw = 0.$$

The variances are then

$$\sigma_t^2 = \int t^2|f(t)|^2 dt \quad \text{and} \quad \sigma_w^2 = \int w^2\left|\hat{f}(w)\right|^2 dw.$$

Then σ_t measures the time duration of the signal $f(t)$ while σ_w measures the frequency spread. The problem is to prove that

$$\sigma_t \sigma_w \geq 1/(4\pi).$$

Prove also that equality holds for $f(t) = C\exp(-at^2)$, with $a > 0$, C chosen to make $\|f\|_2 = 1$. You may *assume* that the functions involved are Schwartz functions. Then use Theorem 1.2.1, part (3) and Exercise 1.2.5, part (c) to see that it is equivalent to show that

$$\int |xf(x)|^2 dx \int |f'(x)|^2 dx \geq \frac{1}{4}.$$

See Dym and McKean [147, p. 119], for the details when f is only square integrable.

1.2.3 Laplace Transforms

Many engineers prefer to use the Laplace transform rather than the Fourier transform.

Definition 1.2.5. The (one-sided) **Laplace transform** $\mathcal{L}$ of a function $f : \mathbb{R}^+ \to \mathbb{C}$ is defined for $s \in \mathbb{C}$ by

$$(\mathcal{L}f)(s) = \int_0^\infty f(t)e^{-st} dt.$$

One advantage is that $\mathcal{L}f$ tends to exist when $\hat{f}$ does not; for example, when f is merely bounded or even exponentially increasing, assuming that the real part of s is sufficiently large. Another advantage is that one can make use of the power of the theory of complex variables (see, e.g., Exercise 1.2.20). It is also possible to define Laplace transforms of functions of several positive real variables and of distributions. See Doetsch [136], Erdélyi et al. [165], Folland [182], Oberhettinger and Badii [502], Schwartz [565], Sneddon [609], or Widder [735] for more information about the Laplace transform. Programs such as Mathematica will compute Laplace transforms.

Exercise 1.2.17 (Properties of the Laplace Transform). We write $\mathcal{L}f$ for $\mathcal{L}f(s)$.

(a) **Existence.** We shall say that *a* function $f:\mathbb{R}^+\to\mathbb{C}$ is **admissible** if it is piecewise continuous on every finite interval *and* if there are constants a in $\mathbb{R}$ and M in $\mathbb{R}^+$ such that

$$|f(x)| \leq M\exp(ax) \quad \text{for all} \quad x \quad \text{in} \quad \mathbb{R}^+.$$

Show that if f is admissible, then $\mathcal{L}f(s)$ exists when $\operatorname{Re} s > a$.

(b) **Derivatives.** Suppose that f and f' are admissible. Then show that

$$\mathcal{L}(f') = s(\mathcal{L}f) - f(0) \quad \text{and} \quad (\mathcal{L}(f))' = \mathcal{L}(-xf(x)).$$

(c) **Integration.** Suppose that f is admissible and that $|f(x)/x| \leq Cx^{\varepsilon-1}$, for $\varepsilon > 0$. Then show that

$$\int_s^\infty \mathcal{L}f(p)dp = \mathcal{L}(f(x)/x) \quad \text{and} \quad \mathcal{L}\left(\int_0^x f(t)dt\right) = \frac{1}{s}\mathcal{L}(f).$$

(d) **Convolution.** Suppose that f, g, *and* fg are all, admissible. Then prove that

$$(\mathcal{L}f)(s)(\mathcal{L}g)(s) = \mathcal{L}\left(\int_0^x f(x-t)g(t)dt\right),$$

$$\mathcal{L}(fg)(p) = \frac{1}{2\pi i}\int_{\mathrm{Re(s)}=c}(\mathcal{L}f)(s)(\mathcal{L}g)(p-s)ds$$

Exercise 1.2.18 (Inversion of the Laplace Transform). Suppose that $e^{-cx}f(x)$ lies in $L^1(\mathbb{R})$ and $f(x)$ vanishes for negative x. Assume also that $f(x)$ is piecewise differentiable. Use Exercise 1.2.7 to show that

$$\frac{1}{2}\left(f(x+)+f(x-)\right) = \lim_{r\to\infty}\frac{1}{2\pi i}\int_{c-ir}^{c+ir} e^{sx}\mathcal{L}f(s)ds.$$

Exercise 1.2.19. Compute the Laplace transform of

(a) x^a, $\operatorname{Re}(a) > -1$.
(b) $\delta(x-a)$, $a > 0$,

(c) the square wave of period a which has the values

$$f(t) = \begin{cases} 1, \, 0 < t < a/2, \\ 0, \, a/2 < t < a. \end{cases}$$

Here $f(t+a) = f(t)$ for all t.

Exercise 1.2.20 (Fourier Series via Laplace Inversion).

(a) If $f(x)$ is periodic, with period $a > 0$; i.e., $f(x+a) = f(x)$, show that

$$(\mathcal{L}f)(s) = (1 - e^{-as})^{-1} \int_0^a f(x) exp(-sx) dx.$$

(b) Use (a) to obtain the Fourier series expression for a periodic function f, assuming that the function $f(x)$ is sufficiently nice.

Hint. *See Sneddon [609, pp. 166–169]. For part (b) use Cauchy's residue theorem to see that the Laplace inversion integral gives a Fourier series via part (a).*

The Laplace transform can be used in the same sort of way as the Fourier transform to solve differential equations. It has also seen much use in the study of the asymptotics of functions. For example, it is useful in number theory where one wants to study asymptotic properties of the sequence of prime numbers. Similarly, as we shall see, it allows one to study the asymptotic properties of eigenvalues of the Laplace operator, by making use of solutions of the heat equation. The Laplace transform results needed here are called Tauberian theorems. Such theorems involve Laplace–Stieltjes transforms:

$$\int_0^\infty \exp(-st) d\alpha(t),$$

with α of bounded variation and *normalized* to make $\alpha(0) = 0$ and $\alpha(x) = (\alpha(x+) + \alpha(x-))/2$. See Apostol [11] for a treatment of Riemann–Stieltjes integrals. Before thinking about Tauberian theorems, however, one should consider the following.

Theorem 1.2.4 (An Abelian Theorem). *Suppose that $f(s) = \int_0^\infty \exp(-st) d\alpha(t)$ for $s > 0$, where α is a normalized function of bounded variation and*

$$\alpha(t) \sim At^c/\Gamma(c+1), \quad \text{as } t \to \infty \quad (or\, t \to 0+).$$

Then

$$f(s) \sim As^{-c}, \quad \text{as } s \to 0+ \quad (or\, s \to \infty).$$

The Abelian theorem is not hard to prove (e.g., see Widder [735]). Its converse is not true without extra hypotheses. Such a converse follows.

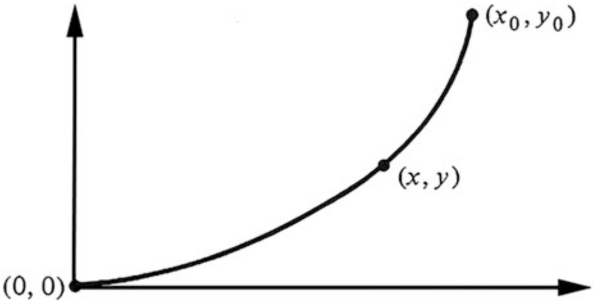


Fig. 1.1 A particle sliding along a curve under the influence of gravity

Theorem 1.2.5 (A Tauberian Theorem). *Suppose that $\alpha(t)$ is a nondecreasing normalized function of bounded variation. And suppose that*

$$f(s) = \int_0^\infty \exp(-st)d\alpha(t)$$

converges for all $s > 0$. If for some nonnegative number c

$$f(s) \sim As^{-c}, \quad \text{as } s \to 0+ \quad (\text{or } s \to \infty),$$

then

$$\alpha(t) \sim At^c/\Gamma(c+1), \quad \text{as } t \to \infty \quad (\text{or } t \to 0+).$$

There is a proof of the preceding theorem in Widder [735, p. 192]. We will see an application of this result in the next section.

Exercise 1.2.21 (Abel's Integral Equation: The Tautochrone). Consider a particle of mass m sliding along a curve under the influence of gravity (with no friction), starting at (x_0, y_0) and ending at $(0,0)$, as pictured in Fig. 1.1. Let $T(y_0)$ be the time for the particle to fall from (x_0, y_0) to $(0,0)$, assuming the shape of the curve is given by some function $y = f(x)$. Then conservation of energy says

$$\sqrt{2g}T(y_0) = \int_{y=0}^{y_0} \varphi'(y)(y_0 - y)^{-1/2}dy, \quad \varphi'(y) = \left(1 + \left(\frac{dx}{dy}\right)^2\right)^{1/2},$$

where g is the acceleration of gravity and the curve is assumed to look like that in Fig. 1.1. Derive this integral equation.

Now assume that $T = T_0 =$ constant; i.e., that the time of descent is independent of the starting point. The tautochrone problem is to find the curve $y = f(x)$ under this hypothesis. Solve this problem by taking the Laplace transform of both sides of the integral equation. Then use the convolution property of the Laplace transform plus the fact that $\mathcal{L}(y^{-1/2}) = (\pi/s)^{1/2}$. The result is that $\varphi'(y) = (2gT_0^2\pi^{-2}y^{-1})^{1/2}$. This gives a differential equation for $x(y)$, which has a cycloid for its solution.

1.2.4 Application to Probability and Statistics: The Central Limit Theorem

The Fourier transform can also be used to derive the central limit theorem, which is one of the most fundamental results in probability and statistics. Special cases of this theorem go back to DeMoivre and Laplace. The general result was proved by Lindeberg in 1922. Previously Liapunov, Chebychev, and Markov had proved the theorem under more restrictive hypotheses. The proof using Fourier transforms was developed from 1925 to 1940 using P. Lévy's proofs of properties of Fourier transforms of probabilistic density functions called characteristic functions. References for probability and statistics are Cramér [115], Feller [177], Kolmogorov [365], and Loève [425]. Collections of papers on applications to physics are Bharucha-Reid [38] and Wax [723].

Let us give a brief summary of the main definitions. For a historical perspective on the subject, see M. Kac [331]. For example, Kac recalls that in 1936 "independent random variables were to me (and others, including my teacher Steinhaus) shadowy and not really well-defined objects."

Let S be a measurable set in some Euclidean space, Let P be a **probability measure** on the Borel sets in S. This means that $P(S) = 1$. A **random variable** $X : S \to \mathbb{R}$ is a measurable function. The **distribution function** F_X of X is defined by

$$F_X(t) = P(\{w \in S \mid X(w) \leq t\}) = P(X \leq t).$$

If

$$F_X(t) - F_X(a) = \int_a^t f_X(u)du$$

for some nonnegative measurable function f_X, we call f_X the **density function** for X. The density function can be obtained via the Radon–Nikodym theorem provided that the probability distribution on $\mathbb{R}$ is absolutely continuous with respect to Lebesgue measure. Otherwise one can use the Lebesgue decomposition theorem to split the probability distribution on $\mathbb{R}$ into a sum of an absolutely continuous and a singular part (see Lang [387], Feller [177, Vol. II, pp. 138–143], or Loève [425]).

Given two random variables X,Y, their **joint distribution** is

$$F(a,b) = P\{w \in S \mid X(w) \leq a \quad \text{and} \quad Y(w) \leq b\} = P(X \leq a,\, Y \leq b).$$

We say that X and Y are **independent** if $F(a,b) = F_X(a)F_Y(b)$.

If X is a random variable, we say that the **expectation** or **mean value** of X is

$$\mu = E(X) = \int_{-\infty}^{+\infty} t f_X(t)dt = \int_{-\infty}^{+\infty} t dF_X(t).$$

The **variance** of X is

$$\sigma^2 = \operatorname{Var}(X) = E((X-\mu)^2) = \int_{-\infty}^{+\infty} (t-\mu)^2 f_X(t)dt,$$

where σ is the **standard deviation**.

Exercise 1.2.22. Suppose that X and Y are independent random variables with density functions f_X and f_Y. Prove that the density function of the random variable $X+Y$ is the convolution $f_X * f_Y$, where convolution is defined by formula (1.2) of Sect. 1.2.

The **characteristic function** φ_X of a random variable X is the Fourier transform

$$\varphi_X(t) = E(e^{itX}) = \int_{-\infty}^{+\infty} e^{itu} f_X(u)du = \widehat{f_X}(-t/2\pi).$$

Theorem 1.2.6 (Properties of Characteristic Functions). *Suppose that X is a random variable with density function $f_X = f$. Then $f \in L^1(\mathbb{R})$, $f \geq 0$, $\int f(x)\,dx = 1$ and the characteristic function of X, which is $\varphi(t) = \hat{f}(-t/2\pi)$, has the following properties.*

(1) The function φ is continuous, $\varphi(0) = 1$, and $|\varphi(t)| \leq 1$ for all t.

(2) The random variable $aX+b$, for $a,b \in \mathbb{R}$, has the characteristic function

$$\varphi_{aX+b}(t) = e^{ibt}\varphi_X(at).$$

(3) If X and Y are two independent random variables, then the characteristic function of the sum $X+Y$ is

$$\varphi_{X+Y}(t) = \varphi_X(t)\varphi_Y(t).$$

*(4) **Fourier inversion.** Suppose that φ is the characteristic function of the random variable X and suppose that $\varphi \in L^1(\mathbb{R})$. Then X has a bounded continuous density function f given by*

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{-itx}\varphi(t)dt.$$

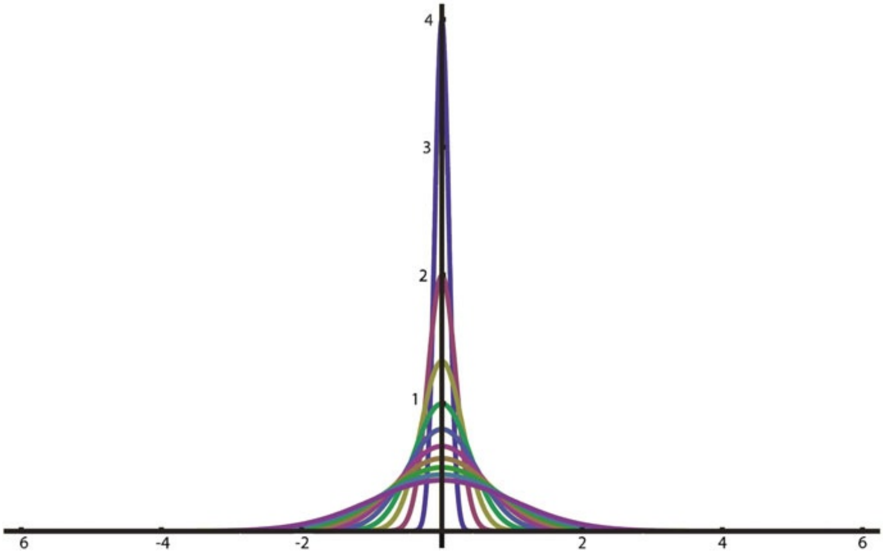


Fig. 1.2 Graphs of normal density curves $G_\sigma(x,0)$, from formula (1.6) with $\sigma = I/10, I = 1,2,\ldots,10$

Exercise 1.2.23. Prove Theorem 1.2.6.

Hint. *See Feller [177, Vol. II, Chap. XV].*

Exercise 1.2.24. Suppose that X and Y are two independent random variables. Show that $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$.

A random variable X is said to *be* **normally distributed or Gaussian with parameters** *m and* σ *or* **normal** (m,σ) if the density function of X is

$$G_\sigma(x,m) = \sigma^{-1}(2\pi)^{-1/2}\exp[-(x-m)^2/(2\sigma^2)]. \tag{1.6}$$

We have graphed some of these density curves in Fig. 1.2.

Exercise 1.2.25 (Properties of Normal Distributions). Assume that X is normal (m,σ); i.e., that its density is $G_\sigma(x,m)$ from formula (1.6).

(a) Show that $E(X) = m$ and $\mathrm{Var}(X) = \sigma^2$.
(b) Prove that

$$P(|X-m| > \lambda\sigma) = \frac{2}{\sqrt{2\pi}}\int_\lambda^\infty \exp(-x^2/2)dx.$$

(c) Show that the characteristic function of X is $\varphi_X(t) = \exp[imt - (\sigma t)^2/2]$.
(d) Show that the sum of any two independent normally distributed random variables is itself normally distributed.

Theorem 1.2.7 (The Central Limit Theorem). *Suppose that* $\{X_n\}_{n\geq 1}$ *is a sequence of independent random variables, each having the same density function* $f(x)$*. Suppose that the mean is* 0 *and the standard deviation is* 1*; i.e.,*

$$\int f(x)dx = 1, \qquad \int xf(x)dx = 0, \qquad \text{and} \qquad \int x^2 f(x)dx = 1.$$

Then $(X_1+\cdots+X_n)/\sqrt{n}$ *is nearly Gaussian for large n; i.e.,*

$$P\left(a \leq \left(\frac{X_1+\cdots+X_n}{\sqrt{n}}\right) \leq b\right) \sim \frac{1}{\sqrt{2\pi}}\int_a^b \exp(-x^2/2)dx, \quad \text{as } n\to\infty.$$

This means that

$$\int_{a\sqrt{n}}^{b\sqrt{n}} (\underbrace{f*\cdots*f}_{\text{n times}})(x)dx \sim \frac{1}{\sqrt{2\pi}}\int_a^b \exp(-x^2/2)dx, \qquad \text{as} \quad n\to\infty.$$

Proof. Using Exercise 1.2.22 and Theorem 1.2.6, one sees that the Fourier transform of the asymptotic relation in the central limit theorem is

$$\hat{f}\left(\frac{s}{\sqrt{n}}\right)^n \sim e^{-2\pi^2 s^2} \quad \text{as} \quad n\to\infty.$$

Recall that $\hat{f}(s) = \int e^{-2\pi isx} f(x)dx$. So the hypotheses of the theorem say that

$$\hat{f}(0) = 1, \ \frac{d}{ds}\hat{f}(0) = 0, \ \frac{d^2}{ds^2}\hat{f}(0) = -4\pi^2.$$

It follows, using the Taylor formula, that since $\hat{f}$ has a continuous second derivative,

$$\hat{f}\left(\frac{s}{\sqrt{n}}\right) = 1 + \frac{(\hat{f})'(0)}{\sqrt{n}}s + \frac{(\hat{f})''(0)}{2n}s^2 + \cdots \sim (1-2\pi^2 s^2/n).$$

Therefore $\hat{f}(s/\sqrt{n})^n \sim e^{-2\pi^2 s^2}$, as $n\to\infty$.

To go from the Fourier transform of the central limit theorem back to the result we want, one must know a "continuity theorem for characteristic functions due to Lévy and Cramér (see Cramér [115, pp. 97–98], and Loève [425]). This is analogous to a Tauberian theorem (see Theorem 1.2.5). Instead we choose to follow the discussion in Dym and McKean [147, p. 116].

We have proved that

$$\begin{aligned}\lim_{n\to\infty} \varphi_{(X_1+\cdots+X_n)/\sqrt{n}}(2\pi s) &= \lim_{n\to\infty}\int_{\mathbb{R}} f_n(x)\exp(2\pi ixs/\sqrt{n})dx \\ &= \exp[-2\pi^2 s^2],\end{aligned}$$

where

$$f_n(x) = (f * \cdots * f)(x)$$

if the number of $f's$ in the convolution is n.

Let d_n be the density for the random variable $(X_1 + \cdots + X_n)/\sqrt{n}$. Then we have for Schwartz functions k:

$$\lim_{n\to\infty} \int_{\mathbb{R}} d_n(x)k(x)dx = \lim_{n\to\infty} \int_{\mathbb{R}} \widehat{d_n}(s)\hat{k}(s)ds$$
$$= \int_{\mathbb{R}} \exp(-2\pi^2 s^2)\hat{k}(s)ds = \int_{\mathbb{R}} (2\pi)^{-1/2}\exp(-x^2/2)k(x)dx,$$

using the Plancherel theorem, dominated convergence, and part (1) of Theorem 1.2.2. The proof is finished using the following Exercise. □

Exercise 1.2.26. Complete the proof of the central limit theorem by approximating

$$\chi_{[a,b]}(x) = \begin{Bmatrix} 0, \ x \notin [a,b] \\ 1, \ x \in [a,b] \end{Bmatrix}$$

by Schwartz functions k.

It is possible to relax the hypotheses in the central limit theorem and one can also obtain error estimates. See the references for the details.

There is also a multivariate central limit theorem for random vectors in $\mathbb{R}^m$. This can be quickly deduced from the case $m = 1$ (see Anderson [7, pp. 74–75]).

In this section, we have seen the Gaussian or normal distribution in two seemingly different contexts—as the fundamental solution of the heat equation (see Example 1.2.1) and as the limiting density function for the normalized sum of a sequence of independent identically distributed random variables. The Gaussian density also appears in a third context closely related to the previous two; namely, in Einstein's work on Brownian motion. The latter is the motion—visible with a microscope—of tiny particles suspended in a liquid. Such motion had been observed by Robert Brown in 1827, as well as many others before him. Einstein proved in 1905 that if S_t denotes the displacement after t minutes of a particle in Brownian motion then S_t has density function

$$(4\pi Dt)^{-1/2}\exp(-x^2/(4Dt)),$$

where D depends in a very explicit way on the temperature and friction coefficient of the medium. See the references mentioned above for more information on Brownian motion. Another interesting reference is Nelson [493].

We will find in Chap. 3 that we can generalize the central limit theorem to the non-Euclidean case, using the non-Euclidean Fourier transform. The proof will be basically the same.

1.3 Fourier Series and the Poisson Summation Formula

> Jétais d'abord parvenu à plusieurs de ces équations par des éliminations très-laborieuses, mais j'emploie maintenant une règle beaucoup plus générale et très-expéditive pour résoudre une fonction arbitraire quelconque en série de sinus ou de cosinus d'arcs multiples. Ces résultats confirment pleinement l'opinion de Daniel Bernoulli.
>
> Les développements dont il s'agit ont donc cela de commun avec les équations aux différences partielles qu'ils peuvent exprimer les propriétés des fonctions entièrement arbitraires et discontinues; c'est pour cela qu'ils se présentent naturellement pour l'intégration de ces dernières équations, et leur application offre des facilitiés singulières dans les questions des lignes ||, mouvements des fluides, de la propagation du son, des vibrations des corps élastiques, et donne un moyen aisé de déterminer les mouvements avec toute la généralité que l'on obtiendrait de l'emploi des fonctions arbitraires. J'en ai fait une application plus particulière à la question de la propagation de la chaleur et l'on parvient ainsi à reconnaître distinctement comment elle se propage par ondes successives dans l'intérieur des corps.[2]
>
> —From Fourier's 1805 draft of his work on the theory of heat propagation in Grattan-Guinness and Ravetz [230, pp. 185–186].

> Als Fourier in einer seiner ersten Arbeiten über die Wärme, welche er der französischen Akademie vorlegte, (21. Dec. 1807) zuerst den Satz aussprach, dass eine ganz willkürlich (graphisch) gegebene Function sich durch eine trigonometrische Reihe ausdrücken lasse, war diese Behauptung dem greisen Lagrange so unerwartet, dass er ihr auf das Entschiedenste entgegentrat. Es soll sich hierüber noch ein Schriftstück im Archiv der Pariser Akademie befinden. Dessenungeachtet verweist Poisson überall, wo er sich der trigonometrischen Reihen zur Darstellung willkürliche Functionen bedient, auf eine Stelle in Lagrange's Arbeiten über die schwingenden Saiten, wo sich diese Darstellungsweise finden soll. Um diese Behauptung, die sich nur aus der bekannten Rivalität zwischen Fourier und Poisson erklären lässt, zu widerlegen, sehen wir uns genöthigt, noch einmal auf die Abhandlung Lagrange's zurückzukommen; denn über jenen Vorgang in der Akademie findet sich nichts veröffentlicht.
>
> … [Here Riemann discusses a formula in Lagrange's work on the vibrating string].
>
> … Hätte Lagrange in dieser Formel *n* unendlich gross werden lassen, so wäre er allerdings zu dem Fourier'schen Resultat gelangt. Wenn man aber seine Abhandlung durchliest, so sieht man, dass er weit davon entfernt ist zu glauben, eine ganz wirkürliche Function lasse sich wirklich durch eine unendliche Sinusreihe darstellen.[3]
>
> —From Riemann [542, pp. 232–233].

---

[2] At first I arrived at several of these equations by very laborious eliminations, but I now use a much more general and expeditious rule to resolve any arbitrary function in a series of sines or cosines of [integral] multiples of angles. These results fully confirm the opinion of Daniel Bernoulli.

As with partial differential equations, the expansions in question can express the properties of entirely arbitrary and discontinuous functions. It is for that reason that they naturally present themselves for the integration of these last equations and their application offers singular facility for questions of parallel lines, fluid motion, sound propagation, vibration of elastic bodies, and gives an easy method for determining the motion in all the generality that can be obtained through the use of arbitrary functions. I have made a more particular application to the question of the propagation of heat and one thus succeeds in recognizing distinctly how it propagates by successive waves in the interior of a body.

[3] Fourier, in his first papers on heat submitted to the French Academy (December 21, 1807) was the first to formulate the theorem that an arbitrarily (graphically) given function could be represented by a trigonometric series. This statement was so unexpected to old Lagrange that he refuted it vehemently. It is claimed that the Archive of the Paris Academy contains such

1.3.1 Fourier Series

Historically the study of Fourier series of periodic functions preceded the study of Fourier integrals. We have reversed the order because it reverses itself naturally in later chapters of this book. Electrical engineers have sometimes argued that one should look solely at Fourier integrals, since nothing in nature is really periodic (see Bracewell [61, Introduction]). However, Fourier series are of great importance, both in pure and applied mathematics. They are so essential that people have designed machines to compute the coefficients of many specialized sorts of functions. In 1903, for example, Michelson had a machine which would find the Fourier coefficients of a sound using vertical rods, each vibrating at a different frequency. Today people use computers and the fast Fourier transform to analyze all sorts of phenomena, as we shall see later in this section. At this point, let us just quote an article on weather prediction that appeared in *Science* (Vol. 220 (1983), p. 40): "Fourier transform methods solved that problem [of making the computer time necessary for the calculations affordable] in the early 1970s, allowing [weather] forecasters to transform the grid mesh into a wave representation, perform the nonlinear calculations, and transform the results back to a grid with no substantial penalty in computing time."

Suppose that $f : \mathbb{R}^m/\mathbb{Z}^m \to \mathbb{C}$. This means that f has period one in each variable. We can also think of f as a function on the unit hypercube $[0,1]^m$ in $\mathbb{R}^m$, with opposite sides identified. We say that the unit hypercube is a **fundamental domain** for $\mathbb{R}^m/\mathbb{Z}^m$. We can also think of $\mathbb{R}/\mathbb{Z}$ as a circle and $\mathbb{R}^m/\mathbb{Z}^m$ as a product of m circles, or a torus. In the case $m = 2$, we get a doughnut.

Definition 1.3.1. The **Fourier series** of $f : \mathbb{R}^m/\mathbb{Z}^m \to \mathbb{C}$ is

$$\sum_{a\in\mathbb{Z}^m} (f, e_a) e_a(x),$$

where $e_a(x) = \exp(2\pi i\, {}^t a x)$, for $x \in \mathbb{R}^m,\ a \in \mathbb{Z}^m$, and the **Fourier coefficients** are

$$(f, e_a) = \int\limits_{[0,1]^m} f(y)\overline{e_a(y)}\, dy.$$

Note that the functions $e_a(x)$, for $a \in \mathbb{Z}^m$, are eigenfunctions of the Laplacian

---

a document. Nevertheless, whenever Poisson uses trigonometric series for the representation of arbitrary functions he refers to a passage in Lagrange's papers on the vibrating string where such a representation is supposed to be found. In order to refute this statement which can only be explained by the well-known rivalry between Fourier and Poisson we have to come back to Lagrange's paper because nothing has been published about Lagrange's opposition in the Academy.

If Lagrange had considered these formulas for infinite n he would have obtained Fourier's result. However, when one goes through his paper one sees that the notion of representing an arbitrary function by an infinite trigonometric series is very foreign to Lagrange.

$$\Delta = \frac{\partial^2}{\partial x_1^2} + \cdots + \frac{\partial^2}{\partial x_m^2}$$

which have period one in each variable. The first two statements of the theorem below say that nice functions f are represented by their Fourier series. This can be interpreted as the spectral resolution of Δ on $\mathbb{R}^m/\mathbb{Z}^m$. This parallels the view of Fourier integrals given by formulas (1.4) and (1.5) of Sect. 1.2 . It is this point of view that we will seek to generalize in Chap. 3, where we replace $\mathbb{Z}^m$ by discrete groups of 2×2 integer matrices.

Theorem 1.3.1 (Properties of Fourier Series).

(1) Suppose that f is in $L^2(\mathbb{R}^m/\mathbb{Z}^m)$. Then f is the L^2- limit of the partial sums of its Fourier series. Moreover, we have the ***Parseval identity****:*

$$\sum_{a\in\mathbb{Z}^m} |(f,e_a)|^2 = \|f\|_2^2.$$

(2) Suppose that $f \in L^1(\mathbb{R}^m/\mathbb{Z}^m)$ and that

$$\sum_{a\in\mathbb{Z}^m} |(f,e_a)| < \infty.$$

Then you can change f on a set of measure 0 to make f continuous on $\mathbb{R}^m/\mathbb{Z}^m$ and such that

$$f(x) = \sum_{a\in\mathbb{Z}^m} (f,e_a)e_a(x).$$

(3) Suppose that $f : \mathbb{R}^m/\mathbb{Z}^m \to \mathbb{C}$ has continuous partial derivatives of all orders less than or equal to k. If $k > m/2$, then the Fourier series of f converges uniformly and absolutely to f.

(4) ***Riemann–Lebesgue Lemma:*** *Suppose that $f \in L^1(\mathbb{R}^m/\mathbb{Z}^m)$. Then (f,e_a) approaches zero, as $\|a\|$ approaches infinity.*

Proof. (1) For example, the Stone–Weierstrass theorem (see Lang [386, p. 148]) says that the exponentials $e_a(x)$, $a \in \mathbb{Z}^m$, form a complete orthonormal set in the Hilbert space $L^2(\mathbb{R}^m/\mathbb{Z}^m)$. You can also deduce this easily from Exercise 1.1.3 of Sect. 1.1. The machinery of Hilbert spaces then completes the proof easily.

(2) The difference of $f(x)$ and the sum of its Fourier series is a function with zero Fourier coefficients and is therefore orthogonal to all trigonometric polynomials and thus to all continuous functions. Since continuous functions are dense in L^1, the difference of f and its Fourier series must be zero except on a set of measure zero.

(3) If we use the notation of formula (1.3) from Sect. 1.2, we have

$$(D^a f, e_b) = (2\pi i b)^a (f, e_b).$$

Then $D^a f \in L^2(\mathbb{R}^m/\mathbb{Z}^m)$ implies that

$$\sum_{|a|=k} \sum_{b\in\mathbb{Z}^m} |(f,e_b)|^2[(2\pi b)^a]^2 < \infty.$$

Now there is a constant $c > 0$ such that

$$\sum_{|a|=k} [(2\pi b)^a]^2 \geq c\|b\|^{2k}.$$

So the Cauchy–Schwarz inequality enables us to compare the series of absolute values of Fourier coefficients and the series $\Sigma\|b\|^{-2k}$, which is an Epstein zeta function. To see that the series $\sum \|b\|^{-2k}$ converges for $k > n/2$ you can proceed by developing a higher-dimensional version of the integral test. See Exercise 1.4.5 of Sect. 1.4. Or use Theorem 1.3.5 of this section.

(4) This is clear for a dense set of L^1 functions; e.g., functions in $L^1 \cap L^2$, or infinitely differentiable functions. □

It follows from part (3) of Theorem 1.3.1 that the speed of convergence of the Fourier series increases with the smoothness of the function. This sort of theorem goes back to Dirichlet in 1829 (see Grattan-Guinness and Ravetz [230, pp. 471–474]). Fourier had not quite managed to show that any piecewise-smooth function can be represented by its Fourier series, taking the average value at jump discontinuities, since we can, in fact, allow the function $f(x)$ to have a jump discontinuity at x, setting

$$f(x+) = \lim_{w\to x, w>x} f(w) \quad \text{and} \quad f(x-) = \lim_{w\to x, w<x} f(x).$$

The convergence of Fourier series of piecewise-smooth functions is discussed in the next exercise.

Exercise 1.3.1 (Jump Discontinuities and Fourier Series).

(a) Suppose that $f(x)$ is a piecewise-continuous function on $[0,1)$. Assume that both one-sided derivatives of f exist at x in $[0,1)$. Define the **Dirichlet kernel** $D_n(x)$ by

$$D_n(x) \;=\; \sum_{|k|\leq n} \exp(2\pi ikx) \;=\; \frac{\sin[\pi(2n+1)x]}{\sin(\pi x)}.$$

Then the **nth partial sum of the Fourier series** of f is

$$(S_n f)(x) = \sum_{|k|\leq n} (f,e_a)e_a(x) = (f*D_n)(x) = \int_0^1 f(y)D_n(x-y)dy.$$

Show that $(f * D_n)(x)$ approaches $\frac{1}{2}(f(x+) + f(x-))$ as n approaches infinity.

(b) Generalize the results of part (a) to functions of several variables. Assume that the function has sufficiently many continuous derivatives. Use the Dirichlet kernel:

$$D_a(x) = \prod_{j=1}^{m} D_{a_j}(x_j).$$

Then $f*D_a$ is the partial sum of the Fourier series of f corresponding to the sum over $b \in \mathbb{Z}^m$ with $|b_j| \le a_j, j = 1, \ldots, m$. Show that $D_a * f$ approaches f as all the a_j approach infinity, assuming that $f(x)$ is sufficiently smooth.

Although we can allow the function $f : \mathbb{R}/\mathbb{Z} \to \mathbb{C}$ to have jump discontinuities, the partial sums of the Fourier series of such functions will always overshoot the mark by about 9 % at the jumps. This is called the **Gibbs phenomenon** because it was pointed out by Gibbs in a letter to *Nature* in 1899. The letter was a reply to Michelson, who was angry that his machine for computing Fourier series failed badly at jumps. Actually the phenomenon was first observed by Wilbraham in 1848. There are many methods for improving the convergence of Fourier series and these are quite important for filter design (see Hamming [247]). The Gibbs phenomenon also occurs for Fourier integrals. See Exercise 2.1.15 of Sect. 2.1.2 for some work on an analogue of the Gibbs phenomenon in a higher-dimensional setting.

Exercise 1.3.2 (The Gibbs Phenomenon and Smoothing in the One-Variable Case).

(a) Set

$$f(x) = \begin{cases} -1, \text{ for} & -\frac{1}{2} \le x < 0 \\ +1, \text{ for} & 0 \le x \le \frac{1}{2}. \end{cases}$$

Show that for large n and small $|x|$, the partial sums $S_n f(x)$ of the Fourier series of f, defined in Exercise 1.3.1, will overshoot the mark by about 9 %. This is the **Gibbs phenomenon**. Either graph the result by computer or follow the theoretical approach in Dym and McKean [147, p. 43].

(b) Show that it is possible to smooth out the horns of the Gibbs phenomenon in part (a) by averaging the first n partial sums (**Cesàro summation**). Show that, in fact,

$$(K_n * f)(x) = \frac{1}{n+1} \sum_{j=0}^{n} S_j f(x),$$

with $K_n =$ the **Fejér kernel** in Example 1.1.1 of Sect. 1.1. Graph the result by computer showing that Fejér smoothing eliminates the oscillations of the partial sums but it has a slow "rise time."

(c) Graph the result of the more popular smoothing method called **Lanczos smoothing**, which is defined by

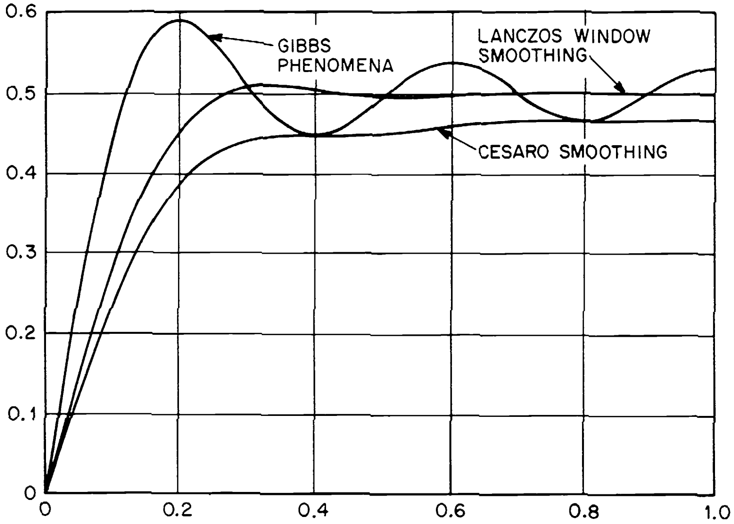


Fig. 1.3 Comparison of various smoothing techniques. (From R.W. Hamming [247, p. 74]. Reprinted by permission of Prentice-Hall, Inc., Englewood Cliffs, NJ)

$$h_n f(x) = \frac{n}{2\pi} \int_{t-\pi/n}^{t+\pi/n} S_n f(u) du.$$

See the Matlab website for information about smoothing and filters or see Hamming [247]. Your answers should resemble the graph of Fig. 1.3.

Carleson [75] proved that the Fourier series of any L^2 function on $\mathbb{R}/\mathbb{Z}$ converges almost everywhere. For $\mathbb{R}^2/\mathbb{Z}^2$, the work of C. Fefferman and others showed that the answers to the convergence question depend on the type of partial sums used; e.g., sums over squares, rectangles, circles (see J. M. Ash [15]). In fact, Ash notes: "At the present time, the passage from 2 to 3 dimensions seems far more substantial and nontrivial than that from 1 to 2 or than that from 3 to more."

There are examples of continuous functions of one variable which have Fourier series that diverge at uncountably many points. And there is an L^1 function with a Fourier series that diverges everywhere (see Kolmogorov [364]). References for such results can be found in the collection of articles edited by J. M. Ash mentioned in the preceding paragraph. See also the work of Zygmund [757]. For some of the older history of Fourier series, see Burkhardt [72], Grattan-Guinness and Ravetz [230], Hilb and Riesz [296], and Riemann [542, pp. 227–271]. Theorem 1.3.4 will show, however, that if all you seek is distributional convergence, then you always get what you want.

The following exercise is discussed more fully in the last section of this chapter.

Exercise 1.3.3 (Weyl's Ergodic Theorem or Wey1's Criterion for Uniform Distribution). For any irrational number a and any continuous function $f(x)$ of period one, show that

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} f(na) = \int_0^1 f(x)dx.$$

Hint. *First prove it for trigonometric polynomials.*

Exercise 1.3.4 (The Shannon Sampling Theorem). This result was proved by Shannon in [587]. A function for which the Fourier transform is zero outside of some interval is called **band limited**. We seek to reconstruct such a function from samples taken at equally spaced points. Suppose that $f : \mathbb{R} \to \mathbb{C}$ and $\hat{f}(y) = 0$ for all y with $|y| > y_c$. Show that

$$f(x) = \sum_{n\in\mathbb{Z}} f\left(\frac{n}{2y_c}\right) \frac{\sin[\pi(2y_c x + n)]}{\pi(2y_c x + n)}.$$

Hint. *First expand $\hat{f}(y)$ in a Fourier series over the interval $[-y_c, +y_c]$:*

$$\hat{f}(y) = \sum_{n\in\mathbb{Z}} A_n \exp\left(\frac{\pi i n y}{y_c}\right) \quad \text{for} \quad y \in [-y_c, +y_c].$$

Let $p_{y_c}(x)$ be 1 if $x \in [-y_c, +y_c]$ and 0 otherwise. Then, for any $y \in \mathbb{R}$,

$$\hat{f}(y) = p_{y_c}(y) \sum_{n\in\mathbb{Z}} A_n \exp\left(\frac{\pi i n y}{y_c}\right).$$

Use the Fourier inversion theorem, the translation property of the Fourier transform, and the formula from Exercise 1.2.6 of Sect. 1.2 giving the Fourier transform of p_{y_c}, and you will find the solution of the problem.

1.3.2 Vibrating Drums

A vibrating drum D is modelled by the wave equation $\Delta u = u_{tt}$ in the interior of D. Assume, for example, that the drum is tied down on the boundary (i.e., u vanishes on the boundary, the Dirichlet boundary condition) and that the initial condition says $u(x,y,0) = f(x,y)$ and $u_t(x,y,0) = 0$. Separation of variables allows you to write the solution as a series involving the eigenfunctions of the Laplacian on the region. For the unit square, the eigenfunctions of the Laplacian have the form $v_{m,n}(x,y) = \sin(\pi nx)\sin(\pi my)$, $n,m = 1,2,3,\ldots$. The corresponding eigenvalues are

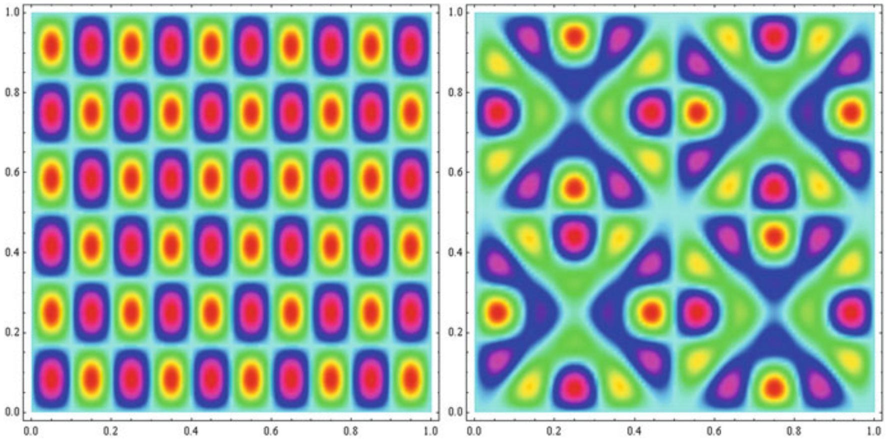


Fig. 1.4 On the *left* is $\sin(10\pi x)\sin(6\pi y)$ and on the *right* $\sin(10\pi x)\sin(6\pi y) - \sqrt{2/3}\sin(6\pi x)\sin(10\pi y)$

$-\pi^2\left(n^2+m^2\right)$. These give the fundamental frequencies of vibration. When you solve the wave equation for the initial conditions just stated, you get a Fourier series:

$$u(x,y,t) = \sum_{n,m\geq 1} c_{n,m}\sin(\pi nx)\sin(\pi my)\cos\left(\pi t\sqrt{n^2+m^2}\right),$$
$$\text{where}\quad f(x,y) = \sum_{n,m\geq 1} c_{n,m}\sin(\pi nx)\sin(\pi my).$$

Thus when you hear a drum you are hearing the eigenvalues of Δ. This led to a famous question of Kac and Bochner: **"Can you hear the shape of a drum?"** It was asked by M. Kac in a famous paper [332]. Mathematically the question becomes: Suppose 2 drums have the same eigenvalues of Δ, does it follow that then you can obtain one drum from the other by rotation and translation? Kac says in his paper that the question was first asked by Bochner. In his autobiography [333, p. 144], he notes that Lorenz conjectured that the area of the drum could be determined by the eigenvalues of Δ (actually for the analogous situation in three dimensions). Weyl proved the Lorenz conjecture in less than 2 years. In 1936 Pleijel showed that the length of the boundary of the drum could be determined from the spectrum of Δ. We will have more to say about hearing the shape of the drum in Chap. 3.

In Fig. 1.4, we imitate Fig. 3 on p. 302 of Courant and Hilbert [111]. Courant and Hilbert graphed the nodal lines of eigenfunctions of Δ on a square. These are graphs of the curves given by setting the eigenfunctions or linear combinations of several of them equal to 0. Such graphs can be seen in real life by placing sand on a drum. Instead we graph density plots for such functions. The graph on the left is that of $\sin(10\pi x)\sin(6\pi y)$ obtained using the Mathematica command:

```
DensityPlot[(Sin[10*Pi*x]*Sin[6*Pi*y]),{x,0,1},{y,0,1},
PlotPoints->100,ColorFunction->Hue,Mesh->False,Axes->False]
```

The graph on the right is that of $\sin(10\pi x)\sin(6\pi y) - \sqrt{2/3}\sin(6\pi x)\sin(10\pi y)$, obtained using a similar Mathematica command.

Exercise 1.3.5. Solve the wave equation on the unit square or unit disc with initial condition modelling of what would happen if the drum were "hit" at a point. Then make a Mathematica movie of your solution. See Courant and Hilbert [111, pp. 300–306]. For the unit disc, one should switch to polar coordinates. The solution involves J-Bessel functions which we discuss in Sect. 2.2. See also Powers [520, pp. 262–269], a book with a cover showing photographs of vibrating kettledrums with dark dust on them—pictures that compare well with the computed nodal lines on p. 269 of the book (which are circles centered at the origin plus rays emanating from the origin). I reproduced the cover of that book in [668, p. 88]. The photographs were themselves taken from a *Scientific American* article by Thomas D. Rossing on the physics of kettledrums, from November, 1982.

Exercise 1.3.6. If you add a forcing term to the wave equation and solve the problem of forced vibration of a membrane, the PDE becomes: $u_{tt} - \Delta u = f(x,y)\cos(\omega t)$. Assume also that $u = 0$ on the boundary of the membrane and that the membrane is at rest with 0 velocity at time 0; i.e., $u(x,y,0) = u_t(x,y,0) = 0$. Then write the solution to the problem as

$$u(x,y,t) = \sum_{n,m\geq 1} c_{n,m}(t)\sin(\pi n x)\sin(\pi m y).$$

Show that then if $v_{n,m}(x,y) = \sin(\pi n x)\sin(\pi m y)$ and $\lambda_{n,m} = -\pi^2(n^2+m^2)$, then

$$c_{n,m}(t) = \frac{(f, v_{n,m})}{|\lambda_{n,m}| - \omega^2}\left(\cos(\omega t) - \cos\left(\sqrt{|\lambda_{n,m}|}t\right)\right).$$

Here

$$(f, v_{n,m}) = \int\limits_{x,y\in[0,1]} f(x,y)v_{n,m}(x,y)\,dxdy.$$

What happens to the solution if ω approaches $\pm\sqrt{|\lambda_{n,m}|}$? This behavior is called "resonance." It leads to a blow-up of the membrane over time.

There are many famous bridge failures due at least partially to resonance. One was Angers Bridge, in Angers, France, 16 April 1850. The collapse was due to resonance from marching soldiers. A similar story holds for the collapse of the Broughton suspension bridge near Manchester, England in 1831. For this reason, soldiers are told not to march in step when going across bridges.

The Tacoma Narrows Bridge self-destructed quite spectacularly in 1940. There are amazing movies of the destruction on the web. Some texts explain the disaster by resonance. Recently that has been disputed. Nevertheless eigenvalues are still important quantities to know in determining the stability of such structures.

> M. Braun [64, pp. 173–175] notes: “There were many humorous and ironic incidents associated with the collapse of the Tacoma Bridge. When the bridge began heaving violently, the authorities notified Professor F. B. Farquharson of the University of Washington. Professor Farquharson had conducted numerous tests on a simulated model of the bridge and had assured everyone of its stability. The professor was the last man on the bridge. Even when the span was tilting more than twenty-eight feet up and down, he was making scientific observations with little or no anticipation of the imminent collapse of the bridge. ”
>
> “A large sign near the bridge approach advertised a local bank with the slogan ‘as safe as the Tacoma Bridge.”’
>
> “After the collapse of the Tacoma Bridge, the governor of the state of Washington made an emotional speech in which he declared ‘We are going to build the exact same bridge, exactly as before.’ Upon hearing this, the noted engineer Von Karman sent a telegram to the governor stating ‘If you build the exact same bridge exactly as before, it will fall into the exact same river exactly as before.”’

The phenomenon of resonance is quite general. It works for all sorts of vibrating objects, even with nonconstant density and tension, and even in higher dimensions. Of course it can be used for good as well as evil. Consider musical instruments, for example. Can they be played for evil? I suppose you might be able to destroy a glass by making it resonate.

Closer to modelling an actual bridge than a vibrating rectangular membrane is a vibrating rectangular plate with free edges. This changes the PDE to $u_{xxxx}+2u_{xxyy}+u_{yyyy}=u_{tt}$. See the article of Gander and Kwok [190] comparing eigenmodes for this fourth-order PDE and the pictures of the vibrating Tacoma Narrows bridge. The article also discusses the Chladni figures which are the nodal lines for eigenvibrations of a square plate.

1.3.3 The Poisson Summation Formula

Now we want to investigate the connection between Fourier series and Fourier integrals. This is the Poisson summation formula. Suppose that $f:\mathbb{R}^m\to\mathbb{C}$ is *a* Schwartz function. Form the periodic function (of period one in each variable),

$$g(x)=\sum_{a\in\mathbb{Z}^m} f(x+a). \tag{1.7}$$

Exercise 1.3.7. Show that the series (1.7) converges uniformly for all $x\in\mathbb{R}^m$, assuming that f is a Schwartz function.

When $m=1$ and the support of f lies inside $(0,1)$, the graphs of f and g may resemble those in Fig. 1.5. If the support of f is larger, the graph of g will be more complicated.

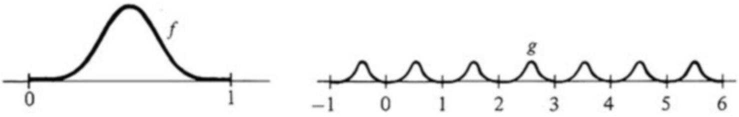


Fig. 1.5 Periodization of a bump function

Theorem 1.3.2 (The Poisson Summation Formula). *If* $f : \mathbb{R}^m \to \mathbb{C}$ *is a Schwartz function, then*

$$g(x) = \sum_{a \in \mathbb{Z}^m} f(x+a) = \sum_{a \in \mathbb{Z}^m} \hat{f}(a) \exp(2\pi i\, {}^t a x).$$

Proof. The Poisson summation formula simply says that the periodic function g is represented by its Fourier series. For $\hat{f}(a)$ is really the ath Fourier coefficient of g, by the following calculation for $a \in \mathbb{Z}^m$:

$$\begin{aligned} \hat{f}(a) &= \int_{\mathbb{R}^m} f(y) \exp(-2\pi i\, {}^t a y) dy \\ &= \int_{[0,1]^m} \sum_{b \in \mathbb{Z}^m} f(y+b) \exp\left[-2\pi i\, {}^t a(y+b)\right] dy \\ &= \int_{[0,1]^m} g(y) \exp(-2\pi i\, {}^t a y) dy. \end{aligned}$$

Thus Theorem 1.3.1 implies Theorem 1.3.2. □

In fact, Poisson is not the only name attached to Theorem 1.3.2, for Gauss and Cauchy also found the formula (see Burkhardt [72, p. 1338]). However, Jacobi, who needed the result in his work on theta functions, attributes it to Poisson (see Jacobi [322, p. 307]).

The hypotheses on the function $f(x)$ in Theorem 1.3.2 should be weakened (see Stein and Weiss [636, pp. 250–257]). There is also an interesting discussion of the Poisson summation formula in Feller [177, Vol. II, p. 632]. Exercise 1.3.22 is a cautionary example on the failure of Poisson summation. See Hejhal [265] for a discussion of a related summation formula due to Voronoi.

It would be useful to be able to derive these results on Fourier series directly from the inversion formula for the Fourier transform in Sect. 1.1. Bracewell ([61, pp. 204 ff.]) claims to do this, but there seems to be a gap in his argument. Exercise 1.2.20 of Sect. 1.2 gives a way of deriving the representation of certain functions by Fourier series, using the inversion of Laplace transforms plus Cauchy's integral theorem (see also Titchmarsh [678, pp. 4–6]). There is also a short distribution-theoretic proof of the Poisson sum formula which we will describe after Theorem 1.3.4. We mention this because it would be nice to be able to create similar arguments in later chapters. There is another distribution-theoretic discussion of the formula in Lighthill [421, pp. 67–68].

On the other hand, it is possible to derive Fourier integrals from Fourier series of functions of period P, by letting P approach infinity. This is done by Titchmarsh in [678, pp. 70–73], for example. To sketch the formal argument, suppose that $f(x)$ is a Schwartz function on $\mathbb{R}$. Let

$$f_P(x) = \sum_{n\in\mathbb{Z}} f(x+nP).$$

The Fourier series of this function f_P of period P is easily seen to be

$$\begin{aligned} f_P(x) &= \frac{1}{P}\sum_{k\in\mathbb{Z}}\left(\int_{-P/2}^{P/2} f(y)\exp(-2\pi iky/P)dy\right)\exp(2\pi ikx/P) \\ &= \frac{1}{P}\sum_{k\in\mathbb{Z}}\int_{-P/2}^{+P/2} f(y)\exp[2\pi i(x-y)k/P]dy. \end{aligned}$$

Then argue that, as P approaches infinity, this last sum approaches the Riemann sum for

$$\int_{w\in\mathbb{R}}\int_{y\in\mathbb{R}} f(y)\exp[2\pi i(x-y)w]dydw.$$

A simpler way of comparing Fourier series with Fourier integrals comes from comparing the kernels that result from approximating the Fourier series by its nth partial sum in Exercise 1.3.1 and approximating the Fourier integral by that on a finite interval $[-r,r]$, as in Exercise 1.2.7 of Sect. 1.2. These kernels are

$$D_n^P(x) = \frac{\sin[\pi(2n+1)x/P]}{P\sin(\pi x/P)}$$

and

$$W_r(x) = \frac{\sin(2\pi rx)}{\pi x}.$$

Now

$$D_n^P(x) = \frac{\sin(\pi 2nx/P)\cos(\pi x/P)}{Psin(\pi x/P)} + \frac{\cos(\pi 2nx/P)\sin(\pi x/P)}{P\sin(\pi x/P)},$$

which approaches $W_{n/P}(x)$ as P approaches infinity (for fixed x).

Next consider Fourier series of periodic distributions. The space of **test functions** $\tilde{\mathcal{D}}$ consists of all infinitely differentiable functions on $\mathbb{R}^m/\mathbb{Z}^m$. Here we should think of $\mathbb{R}^m/\mathbb{Z}^m$ as a product of circles (i.e., as a **torus** which is a doughnut when $m=2$). Convergence of a sequence of test functions means uniform convergence of each sequence of derivatives. A **distribution** T on $\mathbb{R}^m/\mathbb{Z}^m$ is a continuous linear function

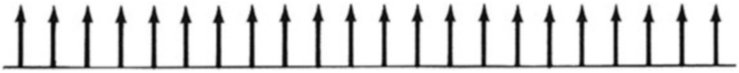


Fig. 1.6 The Dirac comb or Shah functional

$T:\tilde{\mathcal{D}}\to\mathbb{C}$. It is possible to identify T with a periodic distribution $\tilde{T}$ on $\mathbb{R}^m$ with $\tilde{T}_a=\tilde{T}$ for all a in $\mathbb{Z}^m$, using the definition of $\tilde{T}_a$ in Theorem 1.1.1, Part (g) of Sect. 1.1. To see this, suppose that g is a test function on $\mathbb{R}^m$ and set $\tilde{g}=$ the $\mathbb{Z}^m$-**periodization of** g:

$$\tilde{g}(x)=\sum_{a\in\mathbb{Z}^m} g(x+a).$$

Then if T is a distribution on $\mathbb{R}^m/\mathbb{Z}^m$, define the periodic distribution $\tilde{T}$ of period 1 on $\mathbb{R}^m$ by

$$T(\tilde{g})=\tilde{T}(g).$$

For example, if δ is the Dirac delta distribution then

$$\tilde{\delta}=\sum_{n\in\mathbb{Z}^m}\delta_n \tag{1.8}$$

Some engineers call the distribution $\tilde{\delta}$ on $\mathbb{R}^m$ the **shah functional**, for the Russian letter Ш. Note that if f and g are related by formula (1.7), then $\tilde{\delta}f=Шf=g$. One can think of $\tilde{\delta}$ as an infinite "Dirac comb" of impulses pictured in Fig. 1.6 (see the article of Sakai in Vanasse [689, p. 7]). The Fourier transform of δ_a is the function $\exp(2\pi i\,{}^t ax)$. Thus the Fourier transform of the shah functional is

$$\hat{Ш}(x)=\sum_{a\in\mathbb{Z}^m}\exp(-2\pi i\,{}^t ax). \tag{1.9}$$

The Poisson summation formula is exactly the statement that the shah functional is its own Fourier transform; i.e., that $Ш=\hat{Ш}$.

Theorem 1.3.3 (A Criterion for Distributional Convergence of a Trigonometric Series). *A trigonometric series*

$$\sum_{n\in\mathbb{Z}^m} c_n\exp(2\pi i\,{}^t nx)$$

converges in the distributional sense *(see Sect. 1.1) if there are positive constants* C,a *such that*

$$|c_n|\le C\,\|n\|^a,\quad \text{for all}\quad n\in\mathbb{Z}^m.$$

Proof. Set

$$f(x) = \sum_{n \in \mathbb{Z}^m - 0} \|2\pi i n\|^{-2k} c_n \exp(2\pi i\, {}^t n x).$$

This series converges uniformly to a continuous function $f(x)$ provided that k is sufficiently large (by the proof of Theorem 1.3.1). Now apply Δ^k, where Δ is the Laplace operator, to obtain $\Sigma c_n \exp(2\pi i\, {}^t n x)$, which must therefore converge as a distribution. □

Bochner [44] notes that Riemann had already considered this sort of convergence long before Schwartz' book [566].

Definition 1.3.2. Let T be a distribution on $\mathbb{R}^m/\mathbb{Z}^m$. The **Fourier series** of T is $\sum c_n(T) \exp(2\pi i\, {}^t n x)$, with $c_n(T) = \langle T(x), \exp(-2\pi i\, {}^t n x) \rangle$.

Now we can lay to rest all worries about convergence of Fourier series in the sense of distributions.

Theorem 1.3.4. *The Fourier series of any distribution T on $\mathbb{R}^m/\mathbb{Z}^m$ converges to T.*

Proof. First note that the nth Fourier coefficient of $\tilde{\delta}$ on $\mathbb{R}^m/\mathbb{Z}^m$ is 1, since $\tilde{\delta}(e_n) = 1$. And the Fourier series of $\tilde{\delta}$ converges to $\tilde{\delta}$ by Theorem 1.3.1 or Theorem 1.3.2. For this just says that the Fourier series of any test function converges to the function when everything is evaluated at $x = 0$.

Now let T be any distribution on $\mathbb{R}^m/\mathbb{Z}^m$. Then

$$T = T * \tilde{\delta} = T * \left(\sum_{n \in \mathbb{Z}^m} e_n \right) = \sum_{n \in \mathbb{Z}^m} (T * e_n).$$

Here all convolutions are over $\mathbb{R}^m/\mathbb{Z}^m$. Finally,

$$(T * e_n)(x) = \langle T(y), \exp(2\pi i\, {}^t n(x-y)) \rangle = \exp(2\pi i\, {}^t n x) \langle T(y), \exp(-2\pi i\, {}^t n y) \rangle.$$

This completes the proof of Theorem 1.3.4. □

Now let us discuss a distribution-theoretic proof of the Poisson sum formula for $\mathbb{R}/\mathbb{Z}$ from Donoghue [137, p. 162] (cf. also Friedlander [186, p. 104]). Let T be a periodic tempered distribution. Then, since $T_a = T$ for all $a \in \mathbb{Z}$, we have

$$(e^{2nias} - 1)\hat{T}(s) = 0.$$

Thus $\hat{T}(s)$ must be supported on the zeros of $e^{2\pi i a x} - 1$; i.e., on $\mathbb{Z}$. Then, using the first general fact about distributions that was stated in Sect. 1.1, it follows that $\hat{T}$ must be a sum over $a \in \mathbb{Z}$ of linear combinations of derivatives of δ_a. Now suppose that $T = Ш$, the shah functional. Then the fact that shah is its own Fourier transform follows fairly easily.

Next we consider some applications of Poisson's summation formula. The first exercise can also be viewed as a example of the method of images (see Sommerfeld [610, pp. 72–74], and Exercise 1.3.23). We will see many applications of this result throughout this book.

Exercise 1.3.8 (Jacobi's Transformation Formula for the Theta Function). Show that for $x \in \mathbb{R}^m$ and $t > 0$ the following equation is a consequence of Theorem 1.3.2:

$$\sum_{a\in\mathbb{Z}^m} \exp\left[-2\pi i\,{}^t ax - 4\pi^2 c\|a\|^2 t\right] = (4\pi ct)^{-m/2} \sum_{a\in\mathbb{Z}^m} \exp\left[-\|a+x\|^2/(4ct)\right].$$

In the special case $m = 1$, $x = 0$, the theta function of Exercise 1.3.8 is the partition function for the planar rigid rotator in quantum-statistical mechanics. And it approaches the high-temperature limit as $t \to 0+$, a limit that can be computed from the transformation formula in Exercise 1.3.8 (see Hurt [308, p. 10]).

Exercise 1.3.9 (Heat Diffusion on a Circle). Consider the problem

$$\begin{cases} \frac{\partial u}{\partial t} = c\frac{\partial^2 u}{\partial t^2}, & x \in \mathbb{R}/\mathbb{Z},\ t > 0, \\ u(x,0) = f(x), & x \in \mathbb{R}/\mathbb{Z}. \end{cases}$$

Claim. The solution is the same as that of Example 1.2.1 in Sect. 1.2. The exercise is to start out writing the solution as a Fourier series in x. The Fourier coefficients will be functions of t satisfying an ordinary differential equation that is easily solved. This allows you to write the solution $u(x,t)$ as a convolution over $[0,1]$ of f with the theta function in Exercise 1.3.8. Then use that last exercise to see that the solution is really the same as that of Example 1.2.1 in Sect. 1.2.

Exercise 1.3.10. Derive some more results from the Poisson summation formula, such as the following result for $t > 0$:

$$\sum_{n\in\mathbb{Z}} \exp(-|n|t) = \frac{2}{t}\sum_{n\in\mathbb{Z}} \left(1+\left(\frac{2\pi n}{t}\right)^2\right)^{-1}.$$

Justify the formula, even though $\exp(-|x|)$ is not a Schwartz function.

A more general result of this type involves (for $t > 0$)

$$\sum_{n\in\mathbb{Z}} \left(t+n^2\right)^{-s}$$

when $\mathrm{Re}(s) > \frac{1}{2}$. The K-Bessel function will enter into the last formula. And this result will appear in Chap. 3 as the Fourier expansion of Epstein's zeta function.

Theorem 1.3.5 (Asymptotics of the Eigenvalues of Δ on $L^2(\mathbb{R}^m/\mathbb{Z}^m)$). *Let $N(x)$ be the number of eigenvalues λ of Δ such that $|\lambda| \leq x$ and $\Delta h = \lambda h$ for $h \neq 0$ and $h \in L^2(\mathbb{R}^m/\mathbb{Z}^m)$. The eigenvalues are counted with multiplicity. Then*

$$N(x) \ \sim (4\pi)^{-m/2}\Gamma(1+m/2)^{-1}x^{m/2} \quad \text{as} \quad x \to \infty.$$

Proof. Clearly a complete orthonormal set of eigenfunctions of Δ on $\mathbb{R}^m/\mathbb{Z}^m$ consists of the exponentials $e_a(x) = \exp(2\pi i\,{}^t a x),\ a \in \mathbb{Z}^m$. Now

$$\Delta e_a = -4\pi^2 \|a\|^2 e_a.$$

It follows that

$$N(x) = \ \#\{a \in \mathbb{Z}^m \ \big|\ 4\pi^2\|a\|^2 \leq x\}.$$

When $m = 2$ this is the number of lattice points inside of a circle of radius $\sqrt{x}/(2\pi)$. One could give a simple geometrical argument to prove Theorem 1.3.5 (cf. Hardy and Wright [251, pp. 270–272] or Courant and Hilbert [111, p. 430]). But we choose instead to use the Tauberian theorem which was Theorem 1.2.5 of Sect. 1.2.

So we look at the Laplace transform of $N(x)$. This Stieltjes integral is really the sum

$$\sum_{a\in\mathbb{Z}^m} \exp[-4\pi^2\|a\|^2 t] = (4\pi t)^{-m/2} \sum_{n\in\mathbb{Z}^m} \exp[-\|a\|^2/(4t)].$$

Here we have used Exercise 1.3.8, which is a simple application of the Poisson summation formula. The right side of the last equality is asymptotic to $(4\pi t)^{-m/2}$ as $t \to 0+$. Thus the Tauberian Theorem (Theorem 1.2.5 of Sect. 1.2) finishes the proof. $\square$

Finding the order of magnitude of the error term in the asymptotic expression of Theorem 1.3.5 is called the **circle problem** in number theory (when $m = 2$). It appears to be quite difficult. A history of results on the subject can be found in Hejhal [262, p. 450]. Theorem 1.3.5 is a special case of Weyl's result on the asymptotic distribution of the eigenvalues of the Laplacian on a compact domain (see Weyl [730, Vol. I, pp. 393–430, Vol. IV, pp. 432–456]). We will consider a non-Euclidean analogue of Theorem 1.3.5 in Chap. 3.

We take the following exercise from Bogomolny [46, p. 10].

Exercise 1.3.11 (A Trace Formula for Billiards on a Square). The eigenvalues of Δ for periodic boundary conditions on a square in the plane are $\lambda_n = -(2\pi)^2 \|n\|^2$, for $n \in \mathbb{Z}^2$, since periodic eigenfunctions have the form $\exp(2\pi i\,{}^t n x)$, for $x \in \mathbb{R}^2$. Periodic or closed geodesics in the square with sides identified (otherwise known as the torus $\mathbb{R}^2/\mathbb{Z}^2$) are lines of rational slope m_1/m_2, with $m \in \mathbb{Z}^2$. The length of such a geodesic is $\|m\|$. Why? If the fraction is not in reduced form and $d = \gcd(m_1, m_2) > 1$, then we view the geodesic as not what we later call "primitive."

It is the dth power C^d of a primitive geodesic C; i.e., C traversed d times. Suppose that h is a Schwartz function on $\mathbb{R}^+$. Show that

$$\sum_{n\in\mathbb{Z}^2} h\left((2\pi)^2 \|n\|^2\right) = \frac{1}{4\pi}\int_0^\infty h(x)dx + \frac{1}{4\pi}\sum_{m\in\mathbb{Z}^2-0}\int_0^\infty h(x)J_0\left(\|m\|\sqrt{x}\right)\,dx.$$

You can view this as a trace formula with sums over eigenvalues of $-\Delta$ on the torus $\mathbb{R}^2/\mathbb{Z}^2$ on the left and sums over lengths of periodic or closed geodesics in $\mathbb{R}^2/\mathbb{Z}^2$ on the right. These periodic geodesics can be viewed as periodic orbits for the free motion of a billiard on the square billiard table, with periodic behavior at the boundaries. We drew pictures of some periodic geodesics in $\mathbb{R}^2/\mathbb{Z}^2$ in Sect. 1.5.4. See Figs. 1.23 and 1.24.

The Selberg trace formula for $\Gamma\backslash H$, where H is the Poincaré upper half-plane and $\Gamma = SL(2,\mathbb{Z})$ proved in Sect. 3.7 will look rather similar to the result of the last Exercise. In that case, part of the right-hand side of the trace formula can be viewed as the sum over lengths of orbits of non-Euclidean billiards. We drew a picture of such a billiard in Sect. 3.7. See Fig. 3.43.

Next we want to consider an example of Fourier analysis in the real world of smog, earthquakes, etc.

1.3.4 Spectroscopy and the Search for Hidden Periodicities

Spectroscopy was created by Michelson and Morley in 1887 to develop a length standard based on the wavelength of an emission line of an element. In 1892 the wavelength of the red line of cadmium became the international length standard. This lasted until 1960 when the orange line of krypton-86 replaced the red line of cadmium. And Michelson was awarded a Nobel prize in 1907 for this work, which has had an enormous variety of applications—from the identification of air pollutants in Los Angeles smog to the components of the Venusian atmosphere. Fourier analysis was always involved in the theory of spectroscopy. And the theory of spectroscopy has certainly been one of the motivating factors leading to interesting developments in harmonic analysis (see Wiener [737, pp. 119–260]). But Fourier transform spectrometers were not built until the 1970s because they require fast computers and the fast Fourier transform of Cooley and Tukey [108] (see Sect. 1.5.1 and Brigham [67]). The graph of observations of spectra of Venus (Fig. 1.7), from the article of Bell in Vanasse [689, p. 138], shows that Fourier transform spectroscopy has greatly improved resolving power (from 8 cm^{-1} in 1962 to 0.05 cm^{-1} in 1973).

Here we shall only be able to give a brief glimpse into the fascinating theory of spectroscopy. The interested reader could consult any of the following references for more details: Bracewell [62], Bousquet [59], S. P. Davis, M. C. Abrams, and

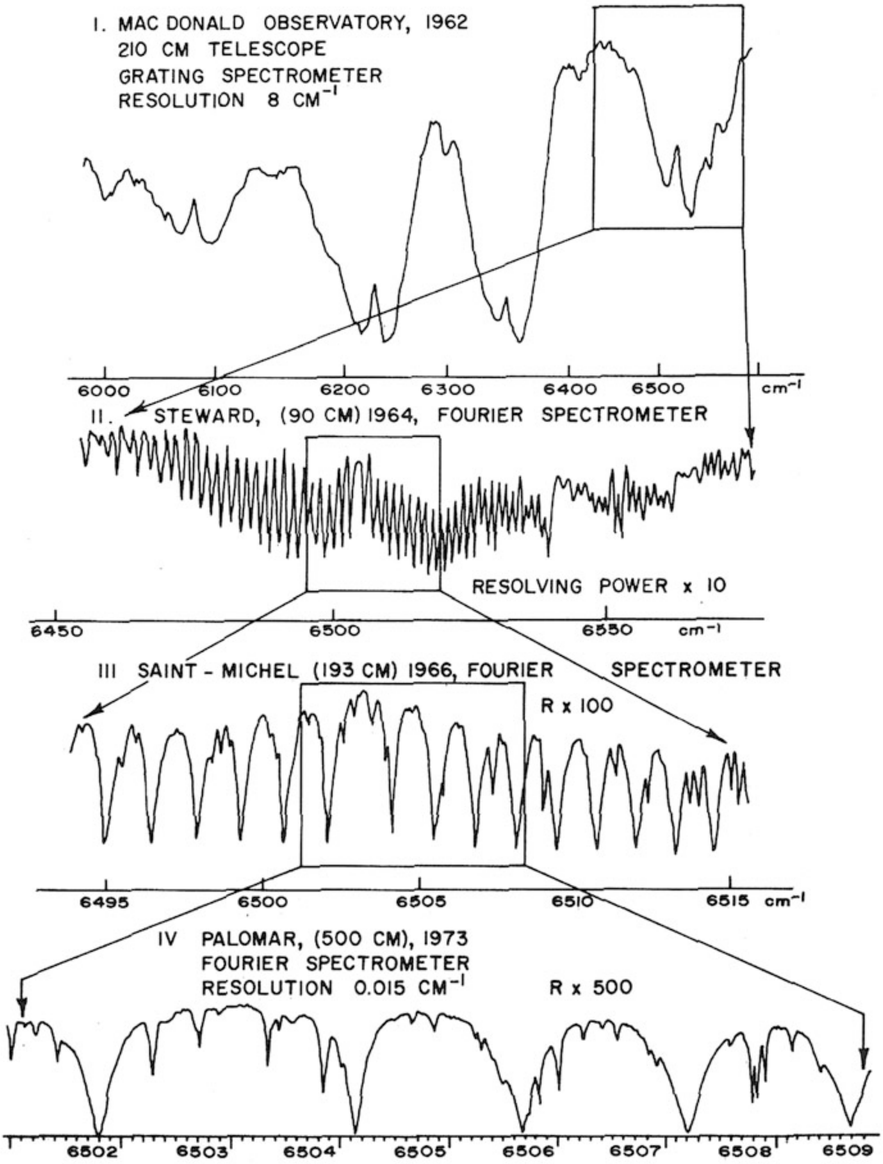


Fig. 1.7 Improvements in the near-IR Venus spectrum due to Fourier spectroscopy: same type detectors (cooled PbS) with almost the same NEP used throughout. Curve I by Kuiper (1962); Curve II from Connes and Connes (1966); Curve III from Larson and Fink (1975); and Curve IV from Connes and Michel (1975). Four strong CO_2 Venusian bands are shown in I; the rotational structure is resolved in II; III shows lines from much weaker overlapping banks; IV gives a good approximation of the true line profile, together with ever fainter line. Trace IV presents approximately $1/800$th of the actual spectral range available from the magnetic tape–general-purpose computer output (parts of which are obscured by H_2O) (after Connes and Michel, 1975). (From Bell's article in Vanasse [689, p. 138]. Reprinted by permission of Academic Press)

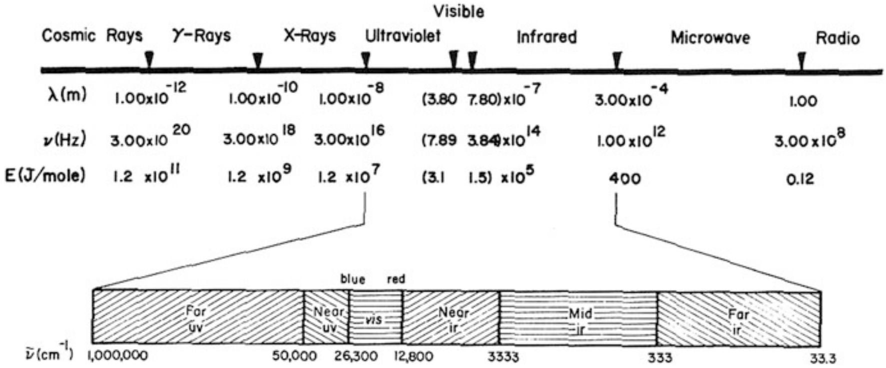


Fig. 1.8 The electromagnetic spectrum. The wavelength λ is given in units of meters; the frequency, ν, is given in units of hertz (1 Hz = 1 oscillation per second); and the energy, E, carried by a **mole** of photons is given in joules (4.184 J = 1 calorie). The wave number $\overline{\nu}$ is expressed in units of cm^{-1} (read "reciprocal centimeters" or "wave numbers"). (From Harris and Bertolucci [255, p. 2]. Reprinted by permission of Oxford University Press)

J. W. Brault [122], Griffiths [232], Harris and Bertolucci [255], Vanasse [689]. Some of the background in crystallography will be discussed in Sect. 1.4. There are also many connections with time-series analysis as Wiener notes in [737, pp. 119–260]. More information on time-series analysis can be found in Bloomfield [42], Jenkins and Watts [328], Kanasewich [336], and Mackey [443, Sect. 18], for example. Time-series analysis studies phenomena $f(t)$ that are neither periodic nor decaying as t approaches infinity. Thus classical Fourier analysis does not apply, although the theory of distributions does allow one to treat such functions. For example, given the variable star data below from Bloomfield [42], one may ask: What are the hidden periodicities? This question goes back to Schuster [563]. We will consider this question briefly after discussing spectroscopy.

Now let us review the basic ideas of spectroscopy. This is the study of the interaction of electromagnetic radiation and matter. Quantum mechanics says that radiation emitted at frequency ν can move an atom from energy level E_1 to E_2 according to Planck's law: $E_2 - E_1 = h\nu$. Here $\nu = c/\lambda_\nu$, where c =the speed of light, λ_ν = wavelength. Since c is known to less accuracy, spectroscopists use $\sigma = 1/\lambda_\nu$ = the **wave number**, and not ν. Figure 1.8 from Harris and Bertolucci [255, p. 2], gives an idea of the wave numbers of various sorts of radiation.

In order to understand what is going on with spectral lines, you really need to know some quantum mechanics, chemistry, crystallography, and group representations. We will touch on this in Sect. 1.4 and Chap. 2.

A Fourier spectrometer uses a Michelson interferometer to divide the beam from some source into two separate beams of equal strength. After the two beams travel different paths, they are recombined. A signal is obtained which ultimately becomes the **interferogram function**

$$F(x) \;=\int_0^\infty B(\sigma)cos(2\pi\sigma x)d\sigma,$$

where $B(\sigma)$ is the **source spectral density** at wave number σ. Then Fourier inversion gives the **equation of spectral recovery:**

$$B(\sigma)=4\int_0^\infty F(x)\cos(2\pi\sigma x)dx.$$

We view $F(x)$ as an even function on the real line here. Since all functions actually live on finite intervals, we have to look at

$$F_T=F\cdot\chi_{[-T,T]},$$

where $\chi_{[-T,T]}$ denotes the indicator function of the set $[-T,T]$; i.e.,

$$\chi_{[-T,T]}(t)=\begin{cases}1, & \text{if} \quad |t|\le T\\ 0, & \text{if} \quad |t|>T.\end{cases}$$

Then the real source spectral density (using the convolution theorem for the Fourier transform and Exercise 1.2.12 of Sect. 1.2) is

$$B_X(\sigma)=B(\sigma)*(\sin(2\pi\sigma X)/\pi\sigma).$$

So spectroscopists call $\sin(2\pi\sigma X)/\pi\sigma$ the **instrument function**. To avoid the Gibbs phenomenon, one may modify the instrument function using methods analogous to those mentioned in Exercise 1.3.2. The spectroscopy literature usually does this by replacing $\chi_{[-X,X]}(t)$ by $\chi_{[-X,X]}(t)(1-|t|/X)$. This replaces B_X by $B*(\sin(\pi\sigma X)/\pi\sigma X)^2$. Physically this "apodization" can be obtained by putting on suitable apertures.

In fact, one can only sample F to reconstruct B and thus Shannon sampling (Exercise 1.3.4) enters the picture. Suppose that the samples of F are taken at nd, $n=0,1,2,3,\ldots$. Then if $B(\sigma)=\hat{F}(\sigma)$ vanishes for $|\sigma|>\sigma_m$, we need $d=1/2\sigma_m$ from the sampling theorem. One says that a sampling distance $d\ge 1/2\sigma_m$ makes the spectrum free of aliasing.

Periodic sampling errors can lead to"satellites" near strong spectral lines, as in Fig. 1.9. Such errors come from defects in the equipment. Let us consider a theoretical explanation.

Figure 1.9. means $B(\sigma)=2C\delta(\sigma-\sigma_0)$, with C a constant. Then $F(x)=C\cos(2\pi\sigma_0 x)$. Suppose that F is sampled with periodic errors. Then x would be replaced by x' with $x'=x+\beta\cos(2\pi\varepsilon x)$. If β is small then

$$F(x')\cong C\left\{\cos(2\pi\sigma_0 x)-\pi\beta\sigma_0\sin[2\pi(\sigma_0+\varepsilon)x]-\pi\beta\sigma_0\sin[2\pi(\sigma_0-\varepsilon)x]\right\}.$$

This gives two satellites as pictured in Fig. 1.9.

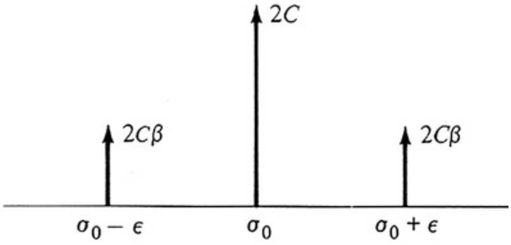


Fig. 1.9 Satellites caused by sampling errors

Finally, let us make a few connections with time-series analysis. The remarks go back to Wiener [737, pp. 119–260]. If $E(t)$ is the electric field and δ the path difference of the interfering beams in the spectrometer, then the **intensity** or **mean power** is, for $\tau = \delta/c$,

$$a_E(\tau) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{+T} \overline{E(t)} E(t+\tau) dt.$$

This expression is called the **autocovariance** or **autocorrelation** of E and plays a large role in time-series analysis. Strictly speaking, it should be normalized to $a(\tau)/a(0)$. It is the Fourier transform of the power spectrum by the **Wiener–Khintchine formula.** To see this formally, suppose $\chi_{[-T,T]}$ is the indicator function of [−T,T] as above. Set $E_T = E \cdot \chi_{[-T,T]}$. Then (under appropriate hypotheses on E)

$$\begin{aligned} a_E(\tau) &= \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{+T} \overline{E(t)} E(t+\tau)\, dt \\ &= \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{+T} \overline{E_T(t)} \int_{-\infty}^{+\infty} \hat{E}_T(w) \exp[2\pi i w(t+\tau)]\, dw\, dt. \\ &= \lim_{T\to\infty} \frac{1}{2T} \int_{-\infty}^{+\infty} \exp(2\pi i w\tau) |\hat{E}_T(w)|^2\, dw. \end{aligned}$$

This is the desired formula since the **power spectrum** is

$$P(w) = \lim_{T\to\infty} |\hat{E}_T(w)|^2/2T.$$

But what does all this have to do with the hidden periodicities in the variable star data of Fig. 1.10 from Bloomfield [42, p. 3]? This is explained quite well by Bloomfield [42, Chap. 2]. The most natural method to try in searching for periodicity is to model the 600 data values x_t, for $t = 0, 1, \ldots, 599$ by the function $(\mu + A\cos(\omega t) + B\sin(\omega t))$ or by a sum of such terms. You could, for example, guess ω, and then choose μ, A, B to minimize

$$S(\mu, A, B) = \sum_{t=0}^{599} (x_t - \mu - A\cos(\omega t) - B\sin(\omega t))^2.$$

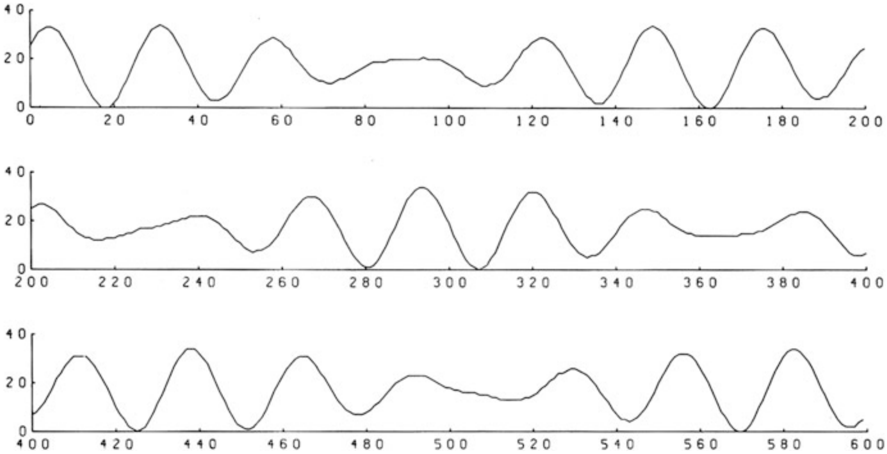


Fig. 1.10 Magnitude of a variable star at midnight on 600 successive nights. (From Bloomfield [42, p. 3]. Reprinted by permission of John Wiley & Sons)

This is the **method of least squares.** If you count the 21 peaks in the data, you could guess the period to be $600/21 \cong 28.6$. This leads you to take $\omega \cong 2\pi/28.6 \cong 0.220$. The method of least squares then gives, if you ignore certain terms,

$$\begin{aligned} \mu &= \bar{x} = \frac{1}{n}(x_0 + \cdots + x_{n-1}),\ n = 599, \\ A(\omega) &= \frac{2}{n}\sum (x_t - \bar{x})\cos(\omega t), \\ B(\omega) &= \frac{2}{n}\sum (x_t - \bar{x})\sin(\omega t). \end{aligned}$$

So we are approximately looking for the discrete Fourier transform.

In order to obtain further information on the periods, one forms the **periodogram** $n(A^2 + B^2)/2$. The graph of this function is shown in Fig. 1.11 (from Bloomfield [42, p. 19]). The maximum occurs at $\omega \cong 0.21644$. If one then does a least squares analysis using this value of ω, one obtains the graph of Fig. 1.12 for $x_t - (\mu + A\ \cos(\omega t)\ + B\ \sin(\omega t))$ (from Bloomfield [42, p. 21]). The graph has an obvious period of 24 days. The new ω is thus about 0.262. Thus one concludes that the original variable star data had two periodic components.

It should be noted that the Shannon sampling theorem affects the choice of ω. If the sampling interval is $d = 1$ day, then $0 \le \omega \le \pi/d$. Every frequency not in the range has an alias in the range.

In Chap. 5 of Bloomfield [42] filtering methods are used to show that x_{300} in the variable star data is in error, with $x_{300} = 18$ and not 19. This rather impressive deduction does not appear to be in the realm of possibility for other examples considered by Bloomfield; e.g., European wheat prices for 1500–1869.

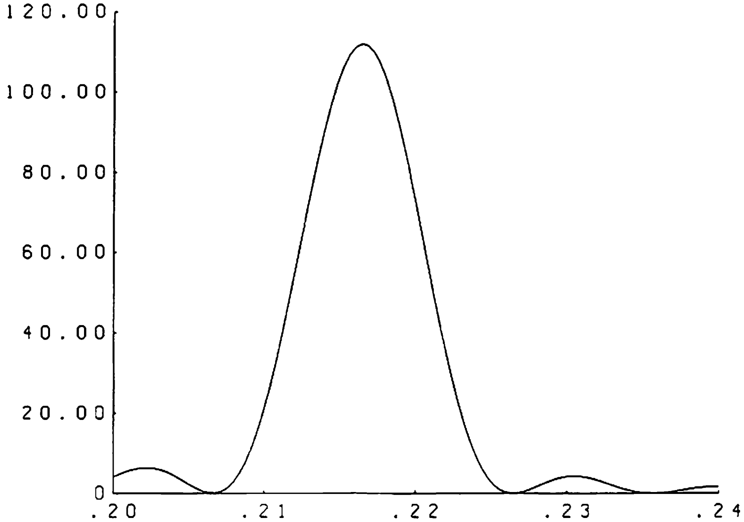


Fig. 1.11 Periodogram of the variable-star data for frequencies ω, $0.20 < \omega < 0.24$. (From Bloomfield [42, p. 19]. Reprinted by permission of John Wiley & Sons)

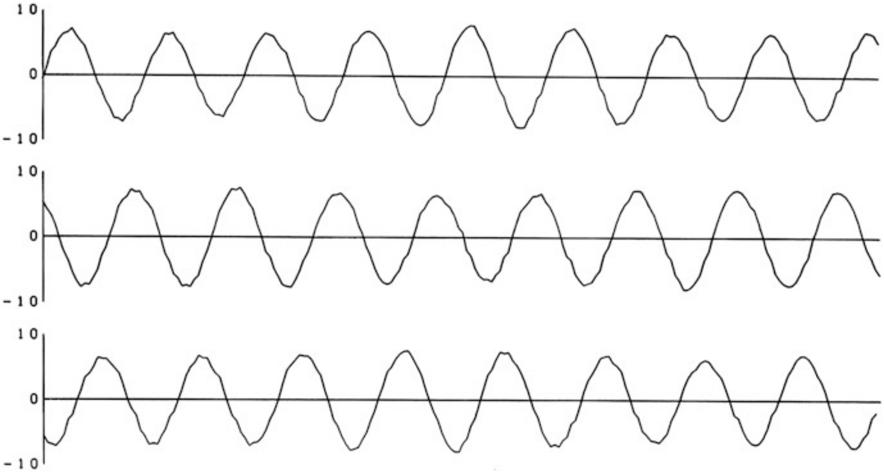


Fig. 1.12 Variable-star data with fitted sinusoid subtracted. (From Bloomfield [42, p. 21]). Reprinted by permission of John Wiley & Sons

1.3.5 Poisson's Sum Formula as a Trace Formula

Our next goal is to obtain an interpretation of Poisson's summation formula in terms of traces of integral operators. In order to do this we shall briefly sketch the theory of integral operators on $L^2(D)$, where D is a measurable subset of $\mathbb{R}^m$. References for this subject include Courant and Hilbert [111], Dieudonné [134], Lang [387], Maurin [458], Stakgold [618, 619] and Yosida [748].

Definition 1.3.3. Let $K \in L^2\left(D \times D\right)$. Then the integral operator K_K defined by

$$L_K f(x) = \int\limits_D K(x,y) f(y) dy \tag{1.10}$$

is called a **Hilbert-Schmidt operator with kernel** K.

Actually, more general Hilbert–Schmidt operators are considered in the references. Clearly L_K maps functions f in $L^2(D)$ to functions $L_K f$ in $L^2(D)$. And just as clearly, the map L_K is linear. So we call it an "**operator**," which just means linear map.

Definition 1.3.4. An operator $L: H \to H$, on a Hilbert space H, is said to be **self-adjoint** if $(Lf,g) = (f,Lg)$, for all f,g in H. Here (f,g) denotes the inner product in the Hilbert space H.

Note that the Hilbert–Schmidt operator (1.10) is self-adjoint if and only if

$$K(x,y) = \overline{K(y,x)}, \text{ for almost all } x,y \in D. \tag{1.11}$$

Definition 1.3.5. An operator $L: H \to H$, on a separable Hilbert space H, is **compact** if for every bounded sequence x_n in H, the sequence Lx_n has a convergent subsequence.

Theorem 1.3.6. *A Hilbert–Schmidt operator on a compact domain is a compact operator.*

For a proof, see Yosida [748, pp. 227–278], or any of the references above, some of which consider only special cases.

Exercise 1.3.12 (Examples of Kernels).

(1) Let $D = (0,1)$ and $K(x,y) = |x-y|^{-a}$, $0 < a < \frac{1}{2}$. Show that this is the kernel of a Hilbert-Schmidt operator on $L^2(D)$. And show that, if instead $\frac{1}{2} \le a < 1$, then the operator is no longer Hilbert–Schmidt, although it is compact.
(2) Show that the kernel $\exp(-x^2 - y^2)$ generates a Hilbert–Schmidt operator on $L^2(\mathbb{R})$, while the kernel $\exp(-|x-y|)$ generates an operator that is not even compact although it is bounded on $L^2(\mathbb{R})$.

Hint. *See Stakgold [619, pp. 353–354].*

Theorem 1.3.7 (The Spectral Theorem for Compact Self-Adjoint Operators). *Suppose that the operator $L : H \to H$ is a compact and self-adjoint on a separable Hilbert space H. Then H has a complete orthonormal set of eigenvectors $\{v_n\}$ of L with $Lv_n = \lambda_n v_n$, $\lambda_n \in \mathbb{R}$, $n = 1, 2, 3, \ldots$. Every vector v in H has a generalized Fourier series representation such that*

$$v = \sum_{n\geq 1} (v, v_n) v_n,$$

$$Lv = \sum_{n\geq 1} \lambda_n (v, v_n) v_n.$$

For a proof of the spectral theorem see the references above. The idea is that one can obtain the eigenvalues of L by finding maxima of the quadratic form (Lv, v) for v in H with $\|v\| = 1$. This method can actually be put on a computer, as the Rayleigh–Ritz and finite element methods. See Strang and Fix [641] for a more complete story of the finite element method.

Exercise 1.3.13 (Expansion of the Kernel of a Self-Adjoint Hilbert–Schmidt Operator). Suppose that $K(x, y)$ is *the* kernel of a self-adjoint Hilbert–Schmidt operator L_K given by (1.10). Let $\{v_n\}$ be a complete orthonormal set of eigenvectors of L_K as provided by Theorem 1.3.7, with $L_K v_n = \lambda_n v_n$. Show that the series

$$\sum_{n\geq 1} \lambda_n v_n(x) \overline{v_n(y)}$$

converges to $K(x, y)$ in the L^2 norm.

If the trace of the self-adjoint integral operator L_K in (1.10) has any meaning, it must be the infinite sum $\sum \lambda_n$ of all the eigenvalues λ_n of L_K. Let us suppose that the operator L_K is **positive**; i.e., that $(L_K f, f)$ is positive for all $f \neq 0$ in H. This means that all the eigenvalues of L_K are positive. Then we have the following theorem.

Theorem 1.3.8 (Mercer's Theorem). *Suppose that L_K is a positive self-adjoint Hilbert-Schmidt operator (1.10) with a continuous kernel K on a compact set in $\mathbb{R}^m$. Then*

$$K(x, y) = \sum \lambda_n v_n(x) \overline{v_n(y)},$$

where the v_n and λ_n are as in Theorem 1.3.7, and the convergence of the series is absolute and uniform.

This is proved in the references.

Exercise 1.3.14 (The Trace of a Positive Self-Adjoint Hilbert-Schmidt Operator). Suppose that the Hilbert-Schmidt operator L_K satisfies the hypotheses of Theorem 1.3.8. Show that

$$\text{Trace } L_K = \sum_{n\geq 1} \lambda_n = \int_D K(x, x) dx.$$

We want to apply this last exercise to deduce Poisson's summation formula. Suppose that $f : \mathbb{R}^m \to \mathbb{C}$ is a Schwartz function. Define an integral operator on functions f in $L^2(\mathbb{R}^m/\mathbb{Z}^m)$ by

$$L_f g(x) = (f * g)(x) = \int_{\mathbb{R}^m} f(x-y)g(y)dy. \tag{1.12}$$

Then the functions $e_n(x) = \exp(2\pi i\, {}^t nx),\ n \in \mathbb{Z}^m$, give a complete orthonormal set of eigenfunctions of L_f on $L^2(\mathbb{R}^m/\mathbb{Z}^m)$. To see this, note that

$$\begin{aligned} L_f e_n(x) &= (f * e_n)(x) = \int_{\mathbb{R}^m} f(x-y)\exp(2\pi i\, {}^t ny)dy \\ &= \int_{\mathbb{R}^m} f(v)\exp(2\pi i\, {}^t n(x-v))dv = \hat{f}(n)e_n(x). \end{aligned} \tag{1.13}$$

On the other hand, we have

$$L_f g(x) = \int_{\mathbb{R}^m/\mathbb{Z}^m} K_f(x,y)g(y)dy,$$

with

$$K_f(x,y) = \sum_{n\in\mathbb{Z}^m} f(x-y-n). \tag{1.14}$$

Now we can use Exercise 1.3.14 to say that

$$\text{Trace } L_f = \sum_{n\in\mathbb{Z}^m} \hat{f}(n) = \int_{\mathbb{R}^m/\mathbb{Z}^m} K_f(x,x)dx. \tag{1.15}$$

But this last integral is easily seen to be $\sum f(n)$, by formula (1.14). Thus (1.15) is really Poisson's summation formula of Theorem 1.3.2. This discussion comes from a seminar talk of Larry Verner in 1973, though it was certainly well known to Selberg as early as 1951. In Chap. 3, we will begin consideration of Selberg's trace formula with an analogous analysis. Our main interest then will be in integral operators on noncompact domains, however.

Exercise 1.3.15 (Band- and Time-Limited Functions—Another Look at Uncertainty). Slepian and Pollak [605] note that Exercise 1.2.16 of Sect. 1.2 does not really tell you "just how close one can come to simultaneously limiting in both time and frequency." Consider f in $L^2(\mathbb{R})$. Set $F(w) = \hat{f}(w/2\pi)$. If $F(w) = 0$ for $|w| > B$, we shall say that f is **band-limited**, and write $f \in \mathcal{B}_B$. Show that the function f in $\mathcal{B}_B$ with $\|f\|_2 = 1$ and

$$\int_{-T}^{+T} |f(t)|^2 dt = \text{maximum}$$

Table 1.2 Values of $\lambda_n(c) = L_n(c) \times 10^{-p_n}$ (from Slepian and Pollack [605, p. 4].)

	$c = 0.5$		$c = 1.0$		$c = 2.0$		$c = 4.0$		$c = 8.0$	
n	L	p	L	p	L	p	L	P	L	p
0	3.0969	1	5.7258	1	8.8056	1	9.9589	1	1.0000	0
1	8.5811	3	6.2791	2	3.5564	1	9.1211	1	9.9988	1
2	3.9175	5	1.2375	3	3.5868	2	5.1905	1	9.9700	1
3	7.2114	8	9.2010	6	1.1522	3	1.1021	1	9.6055	1
4	7.2714	11	3.7179	8	1.8882	5	8.8279	3	7.4790	1
5	4.6378	14	9.4914	11	1.9359	7	3.8129	4	3.2028	1
6	2.0413	17	1.6716	13	1.3661	9	1.0951	5	6.0784	2
7	6.5766	21	2.1544	16	7.0489	12	2.2786	7	6.1263	3
8	1.6183	24	2.1207	19	2.7768	14	3.6066	9	4.1825	4

is the eigenfunction corresponding to the largest eigenvalue of the integral operator

$$Lf(t) = \int_{-T}^{T} f(s) \frac{\sin[B(t-s)]}{\pi(t-s)} ds.$$

Hint. *Consider the operators*

$$\mathcal{B}f(t) = \frac{1}{2\pi} \int_{-B}^{+B} \hat{f}(w/2\pi) \exp(itw) dw,$$

and

$$\mathcal{T}f(t) = \begin{cases} f(t), & \text{if} \quad |t| \le T, \\ 0, & \text{if} \quad |t| > T. \end{cases}$$

Show that $\mathcal{B}\mathcal{T}f = Lf$.

Note. Slepian and Pollak calculate the eigenvalues of L by showing that the corresponding eigenfunctions must be spheroidal wave functions. These results have been generalized to some symmetric spaces by Grünbaum et al. [234]. Slepian and Pollak give a table of values of $\lambda_n(c)$ (Table 1.2) where $c = BT/2$.

Some of the most interesting applications of the spectral theorem for integral operators are in partial differential equations. This occurs because one can often write the inverse operator of a differential operator D (or of $D - uI$, $I =$identity, $u \in \mathbb{C}$) as an integral operator. The kernel of this integral operator is called the **Green's function**. For example, the Green's function $G(x,y)$ for $-\Delta$ on a region D in $\mathbb{R}^3$ with the (Dirichlet) boundary condition that f vanish on the boundary ∂D satisfies

$$\left.\begin{array}{l} (Tf)(x) = \int_D G(x,y) f(y) dy, \\ -\Delta T f = f, \quad \text{for all } f \text{ in } L^2(D), \\ -T\Delta f = f \quad \text{and } f \text{ such that both } f \text{ and } \Delta f \text{ are in } L^2(D) \\ \text{and such that } f \text{ vanishes on the boundary } \partial D. \end{array}\right\} \tag{1.16}$$

Equations (1.16) say that

$$\begin{aligned} -\Delta_x G(x,y) &= \delta(x-y), \text{ for } x,y \text{ in } D \\ G(x,y) &= 0, \ \text{ if } x \text{ is in } \partial D. \end{aligned} \tag{1.17}$$

Thus Exercise 1.1.6 of Sect. 1.1 implies that in 3 dimensions

$$G(x,y) = (4\pi \|x-y\|)^{-1} + h(x,y),$$

where $h(x,y)$ has no singularities and is chosen to make $G(x,y)$ vanish if x lies on the boundary of D. Finding $G(x,y)$ explicitly is difficult or impossible unless the region D is nice. For example, sharp spikes are not allowed in the region D. The following exercise gives a simple example of a Green's function for the Laplace operator on a ball of radius r. You can find more examples and discussions of Green's functions in Sect. 2.2 and in the following references: Courant and Hilbert [111], Dunford and Schwartz [145, Vol. II, Chap. XIII] (where Green's functions are called resolvent kernels), Garabedian [197], Morse and Feshbach [479, Vol. II, Chap. 10], and Stakgold [618, 619]. We will have more to say about Green's functions in Chaps. 2 and 3.

Exercise 1.3.16 (Green's Function for $-\Delta$ on a Ball of Radius r by the Method of Images). William Thompson developed this method and applied it to problems in electro- and magneto-statics. The idea is to use the map of $\mathbb{R}^3$ which sends a point $x \in \mathbb{R}^3$ to its **inverse point** x^* with respect to the sphere of radius r. By this, we mean that x^* lies on the same radial line from the origin as x and $\|x\|\,\|x^*\| = 1$. The problem is to show that the Green's function for the ball of radius r in $\mathbb{R}^3$ is

$$G(x,y) = (4\pi)^{-1}(\|x-y\|^{-1} - r\|y\|^{-1}\|x-y^*\|^{-1}).$$

Hint. *To show that* $G(x,y) = 0$ *when* $\|x\| = r$, *you must use the fact that*

$$\frac{\|x-y^*\|}{\|x-y\|} = \frac{r}{\|y\|} \quad \text{if} \quad \|x\| = r.$$

If the Green's function $G(x,y)$ exists for the region D, then the operator T in formula (1.16) will be a Hilbert–Schmidt operator which means that both T and $-\Delta$ will have a complete orthonormal set of eigenfunctions v_n in $L^2(D)$ with $-\Delta v_n = \lambda_n v_n$ and

$$G(x,y) = \sum_{n \geq 1} \lambda_n^{-1} v_n(x)\overline{v_n(y)}. \tag{1.18}$$

In general, the convergence of this series is only in the L^2 norm. This gives another method for finding the Green's function. However, even in the simple example below, one finds a series which does not converge absolutely.

Exercise 1.3.17 (Green's Function for $-\Delta$on a Rectangular Solid). Consider the Dirichlet eigenvalue problem

$$\begin{cases} -\Delta v = \lambda v & 0 < x < a,\ 0 < y < b,\ 0 < z < c \\ v = 0 & \text{on the boundary of the rectangular solid.} \end{cases}$$

Find a complete orthonormal set of eigenfunctions and consider formula (1.18) for the Green's function. Show that the series does not converge absolutely, using information about Epstein's zeta function (Theorem 1.4.1 of Sect. 1.4). You can find a discussion of the method of images for this problem in Courant and Hilbert [111, pp. 378–384].

1.3.6 Schrödinger Eigenvalues

As a last topic in this discussion of spectral theory, we want to consider a **Schrödinger eigenvalue problem in one dimension**:

$$\left.\begin{array}{c} y''(x) - 4\pi^2 x^2 y(x) = \lambda y(x), \quad x \in \mathbb{R} \\ y(x) \in L^2(\mathbb{R}). \end{array}\right\} \tag{1.19}$$

This is a good, though simplistic, model of a quantum-mechanical problem (see Messiah [465, Vol. I, Chap. XII]), which indicates quite clearly that a continuous problem can sometimes have a only a discrete sequence of possible eigenvalues. In quantum theory this is thought of as saying that the system is quantized and can only be in a certain discrete list of states. The eigenvalues of the differential operator in (1.19) correspond to the energy levels of the system, once everything has been normalized in the proper way for physics. We have chosen our normalization to connect with the theory of Fourier integrals. If the problem were on a finite interval, then we could find a Green's function and use Hilbert–Schmidt theory to see that the spectrum is discrete. However, this problem is singular, because it is on an infinite interval. Thus the spectrum has no a priori reason to be discrete. And certainly there are other Schrödinger eigenvalue problems with mixtures of continuous and discrete spectra. An example is the eigenvalue problem associated with the hydrogen atom which is considered in Sect. 2.1. Another such eigenvalue problem will be discussed in Chap. 3.

We could approach the eigenvalue problem (1.19) by substituting $y(x) = w(x)\exp(-\pi x^2)$. The resulting differential equation for $w(x)$ can be seen to have a polynomial solution only for special values of λ. However, we shall take a different approach—**the factorization method of Infeld and Hull** [316] (see Talman [655]). Set $D = d/dx$. Then we can factor the operator

$$H = D^2 - (2\pi x)^2 = (D - 2\pi x)(D + 2\pi x) - 2\pi = (D + 2\pi x)(D - 2\pi x) + 2\pi.$$

Define the raising operator $A = D - 2\pi x$ and the lowering operator $B = D + 2\pi x$. Then $H = AB - 2\pi = BA + 2\pi$.

Exercise 1.3.18 (Hermite Functions Are Eigenfunctions of H and the Fourier Transform).

(1) Show that if we define $y_0 = \exp(-\pi x^2)$, $y_n = Ay_{n-1}$, then $Hy_n = \lambda_n y_n$, where the eigenvalues are $\lambda_n = -(4n+2)\pi$.
(2) Show that if $\hat{f} =$ the Fourier transform of f, then $\widehat{y_n} = (-i)^n y_n$.
(3) Show that $y_n(x) = (D - 2\pi x)^n \exp(-\pi x^2) = \exp(-\pi x^2) D^n \exp(-2\pi x^2)$.
(4) Show that $y_n(x)$ is the product of $\exp(-\pi x^2)$ and a polynomial of degree n.

Exercise 1.3.19 (Spectral Decomposition of the Schrödinger Operator H). Set $v_n = \|y_n\|_2^{-1/2} y_n$, with y_n as in Exercise 1.3.18. Show that v_n gives a complete orthonormal set for $L^2(\mathbb{R})$ and thus the spectral decomposition of the differential operator $H = D^2 - (2\pi x)^2$.

Hint. *The functions v_n are pairwise orthogonal since they are eigenfunctions of the self-adjoint operator H corresponding to distinct eigenvalues. To see that they form a complete orthonormal set, you must show that if $f \in L^2(\mathbb{R})$ and $(f, v_n) = 0$ for all n, then f is zero (almost everywhere). One way to do this is to look at*

$$\left(f(y)\exp(-\pi y^2)\right)\widehat{\ }\,(x) = \int f(y)\exp(-2\pi ixy - \pi y^2)dy$$

$$= \sum_{n \geq 0} \frac{(-2\pi ix)^n}{n!} \int y^n f(y) \exp(-\pi y^2) dy.$$

If $(f, v_n) = 0$ for all n, this must vanish. Why?

Exercise 1.3.20 (Connection with the Classical Hermite Polynomials). Define

$$H_n(x) = (-1)^n \exp(x^2) D^n \exp(-x^2), \text{ where } D = d/dx.$$

The $H_n(x)$ are the classical Hermite polynomials (see Courant and Hilbert [111, pp. 91–97]). Find the formula connecting y_n and H_n.

Exercise 1.3.21 (A New Expression for the Fourier Transform on $L^2(\mathbb{R})$). This idea comes from Wiener [736, pp. 51–71]. Suppose that the functions v_n are as in Exercise 1.3.19. Then any function f in $L^2(\mathbb{R})$ has an expression $f = \sum (f, v_n) v_n$. What is the corresponding expansion for the Fourier transform of f ?

Exercise 1.3.22. Find an example of $f \in L^1(\mathbb{R})$ such that $\hat{f} \in L^1(\mathbb{R})$, and $f(n) = 0$ for all $n \in \mathbb{Z}$; $\hat{f}(0) = 1$, $\hat{f}(n) = 0$, if $n \in \mathbb{Z}$, $n \neq 0$. This shows that the Poisson summation formula is not always valid when you might expect it to be.

Hint. *See Katznelson [347, Exercise 15, pp. 130–131].*

Exercise 1.3.23 (Periodic Green's Function for the Helmholtz Equation in $\mathbb{R}^3$ by the Method of Images).

(1) Show that $g_k(x,y) = \exp(ik\|x-y\|)/(4\pi\|x-y\|)$ is a Green's function for the Helmholtz operator $\left(\Delta - k^2\right)$ on $\mathbb{R}^3$. Other choices for $g_k(x,y)$ are

$$\exp(-ik\|x-y\|)/(4\pi\|x-y\|),\ \cos(k\|x-y\|)/(4\pi\|x-y\|),$$

etc. Our choice of $g_k(x,y)$ dies off exponentially at infinity, if $\operatorname{Im} k > 0$. It can be viewed as corresponding to outgoing waves in the wave equation.

(2) Use the method of images to obtain the Green's function for $\left(\Delta - k^2\right)$ on the unit cube $\mathbb{R}^3/\mathbb{Z}^3$ corresponding to periodic boundary conditions.

Answer.

$$G_k(x,y) = \frac{1}{4\pi}\sum_{n\in\mathbb{Z}^3}\frac{\exp(ik\|x+n-y\|)}{\|x+n-y\|}.$$

For what values of k does this series converge?

(3) Show that, as in formula (1.18), we can represent the Green's function in part (2) by the following series converging in the L^2-sense:

$$G_k(x,y) = \sum_{n\in\mathbb{Z}^3}\frac{\exp(2\pi i\,{}^t n(x-y))}{\|2\pi n\|^2 - k^2}.$$

For absolute convergence, show that one can look at $G_k - G_m$, since if $R_k = \left(\Delta - k^2\right)^{-1}$, then $R_k - R_m = \left(k^2 - m^2\right)R_k R_m$.

(4) What does Poisson summation do to the formulas in parts (2) and (3)?

1.4 Mellin Transforms, Epstein and Dedekind Zeta Functions

It is remarkable that the deepest ideas of number theory reveal a far-reaching resemblance to the ideas of modern theoretical physics. Like quantum mechanics, the theory of numbers furnishes completely nonobvious patterns of relationship between the continuous and the discrete (the technique of Dirichlet series and trigonometric sums, p-adic numbers, nonarchimedean analysis) and emphasizes the role of hidden symmetries (class field theory, which describes the relationship between prime numbers and the Galois groups of algebraic number fields). One would like to hope that this resemblance is no accident, and that we are already hearing new words about the World in which we live, but we do not yet understand their meaning.

—From Yu. I. Manin, as translated by Ann and Neal Koblitz, *Mathematics and Physics*, Birkhäuser, Boston, 1981.

1.4.1 Mellin Transforms

In this section we shall consider a method used both by number theorists and physicists to obtain analytic continuations of Dirichlet series. Riemann used this procedure in [542, pp. 145ff], to obtain the analytic continuation of the **Riemann zeta function**:

$$\zeta(s) = \sum_{n\geq 1} n^{-s}, \quad \operatorname{Re} s > 1 \tag{1.20}$$

to the entire complex s-plane. Ewald used the same method to compute potentials of crystal lattices (see Born and Huang [57, p. 389]). Riemann's method can also be used to compute Green's functions for a rectangular parallelepiped (see Courant and Hilbert [111, pp. 378–384], and Exercise 1.3.23 of Sect. 1.3). We shall discuss these examples in this section, as well as the Dedekind zeta function of an algebraic number field.

Before considering Riemann's method we must understand another important integral transform.

The **Mellin transform** of $f:\mathbb{R}^+\to\mathbb{C}$ is

$$Mf(s) = \int_0^\infty f(y)y^{s-1}dy.$$

The Mellin transform is as well suited to the study of the multiplicative properties of numbers as the Fourier transform is well suited to the study of the additive properties of numbers. In fact, the change of variables $y = e^x$ allows you to go from one transform to the other. You could also consider the Mellin transform as a two-sided Laplace transform.

Exercise 1.4.1 (Inversion of the Mellin Transform). Make the change of variables $y = e^x$ to see that $Mf(s) = \hat{F}(t/2\pi)$, where $s = \sigma + it$, with σ, t real, and $F(x) = f(e^x)e^{\sigma x}$. Suppose that $F(x) \in L^1(\mathbb{R})$ and that $f(y)$ is piecewise continuous with one-sided derivatives always existing. Use Exercise 1.2.7 of Sect. 1.2 to show that

$$\frac{1}{2}\left(f(y+) + f(y-)\right) = \frac{1}{2\pi i}\lim_{r\to\infty}\int_{c-ir}^{c+ir} y^{-s}Mf(s)ds.$$

Exercise 1.4.2 (Properties of the Mellin Transform). Prove the following properties, assuming that the functions that are Mellin transformed satisfy the conditions of Exercise 1.4.1, for example.

(a) Set $D = yd/dy$. Then $MDf(s) = -sMf(s)$.
(b) $\frac{d}{ds}Mf(s) = M(f(y)\log y)$.
(c) $M(fg)(s) = \frac{1}{2\pi i}\int_{\operatorname{Re} z=c} Mf(z)Mg(s-z)dz$.
(d) For $f, g : \mathbb{R}^+ \to \mathbb{C}$, define $(f * g)(y) = \int_0^\infty f(y/u)g(u)u^{-1}du$. Then

$$M(f*g)(s) = Mf(s)Mg(s).$$

(e) Set $f_a(y) = f(ay)$, for a and y in $\mathbb{R}^+$. Then $M(f_a)(s) = a^{-s}Mf(s)$.
(f) $\int_{\operatorname{Re} s=c} Mf(s)g(s)ds = \int_{y=0}^\infty f(y)M^{-1}g(y)y^{-1}dy$.

Table 1.3 A short table of Mellin transforms

$f(y)$, $y > 0$	$Mf(s)$
$\exp(-ay), \quad y > 0$	$a^{-s}\Gamma(s)$, $\mathrm{Re}\ s > 0$
$\begin{cases} \exp(-y), & y > a \\ 0, & 0 < y < a \end{cases}$	$\Gamma(s,a)$
$\exp\left[\frac{-a}{2}\left(y+\frac{1}{y}\right)\right]$, $a > 0$	$2K_s(a)$
$\exp(-zy)(1+y)^{a-1}$, $\mathrm{Re}\ z > 0$	$\Gamma(s)\Psi(s,a+s,z)$, $\mathrm{Re}\ s > 0$

Exercise 1.4.3. Verify Table 1.3.

Hint. *Here you need to know a little about special functions such as gamma, incomplete gamma, K-Bessel, and confluent hypergeometric functions. We will have much more to say about these functions as we continue our discussion. You can view gamma and incomplete gamma as being defined by the first two lines of this table. See Theorem 1.4.1 for an application of* $\Gamma(s,a)$. *The K-Bessel function* $K_s(a)$ *is discussed in Sect. 3.2. Confluent hypergeometric functions will appear in Volume II [667]. See, for example, Lebedev [401, Chaps. 1, 5, 9]. A longer table of Mellin transforms can be found in Erdélyi et al. [165] or Oberhettinger [498].*

More information on Mellin transforms can be found in Sneddon [609] and Titchmarsh [678]. You could, of course, also consider Mellin transforms of distributions.

It is possible to view the Mellin inversion formula in Exercise 1.4.1 as the **spectral resolution of the singular differential operator** $(yd/dy)^2$ acting on functions in the space

$$L^2(\mathbb{R}^+, y^{-1}dy) = \left\{ f : \mathbb{R}^+ \to \mathbb{C} \quad \text{measurable} \;\middle|\; \int_0^\infty |f(y)|^2 y^{-1} dy < \infty \right\}.$$

The functions y^s are all eigenfunctions of this differential operator. See Stakgold [618, pp. 465–466], for an exercise on the Mellin inversion formula from this point of view. In Chap. 3 and Volume II we will generalize the Mellin inversion formula to spaces of positive $n \times n$ matrices. This is, in fact, one of our main objectives in writing these volumes. Note that the operator $(yd/dy)^2$ and the measure $y^{-1}dy$ are invariant under the change of variables $w = ay$ for $a \in \mathbb{R}^+$. Thus we are really viewing a spectral decomposition that is intimately related to the multiplicative group of positive real numbers when we study the Mellin inversion formula. The generalizations of Mellin inversion to be considered in Chap. 3 and Volume II are directly related to the general linear group of $n \times n$ nonsingular real matrices.

1.4.2 Epstein's Zeta Function

All of the analytic continuations of zeta functions that we want to consider are related to the analytic continuation of Epstein's zeta function. The latter lives on the symmetric space $\mathcal{P}_n$ of positive definite symmetric $n \times n$ real matrices and it will appear in the study of harmonic analysis on $\mathcal{P}_n$ in Chap. 3 and Volume II [667].

Definition 1.4.1. Epstein's zeta function of $Y \in \mathcal{P}_n$ and $s \in \mathbb{C}$, with $\operatorname{Re} s > n/2$ is

$$Z(Y,s) = \frac{1}{2} \sum_{a \in \mathbb{Z}^n - 0} Y\,[a]^{-s}.$$

Here $Y\,[a] = {}^t aYa$, thinking of a as a column vector, with ${}^t a$ =transpose of a. Thus $Y\,[a]$ is the quadratic form

$$Y\,[a] = \sum_{i,j=1}^{n} y_{ij} a_i a_j, \text{ if } Y = (y_{ij}),\ \ a = {}^t(a_1, \ldots, a_n).$$

In the special case that $n = 1$, then $Y = y \in \mathbb{R}^+$, and $Z(Y,s) = y^{-s}\zeta(2s)$, where ζ=Riemann's zeta function. See Edwards [149] for a fascinating treatment of the work on that zeta function, whose main importance for number theory comes from the following formula:

$$\zeta(s) = \prod_{p=\text{prime}} \left(1 - p^{-s}\right)^{-1}, \quad \operatorname{Re} s > 1. \tag{1.21}$$

Exercise 1.4.4. Prove formula (1.21) which is called the **Euler product** for Riemann's zeta function. You need to know that every natural number $n \in \mathbb{Z}^+$ can be factored uniquely (up to order) as a product of primes.

Many other special cases of Epstein's zeta function arose in number theory before Epstein's papers which appeared in 1903 and 1907 (see Epstein [163]). Most of these cases arise from the fact that the Dedekind zeta function of an algebraic number field can be written as a finite sum of integrals of Epstein zeta functions (Theorem 1.4.2). The sad story of the end of Epstein's life and the horrors of Hitler's Germany can be found in Siegel [600, Vol. III, pp. 464–470], or the translation [601].

There are many applications of Epstein zeta functions in statistical and solid-state physics. We will discuss some of these at the end of this section. An interesting reference that gives applications of Epstein's zeta function in quantum-statistical mechanics is Hurt and Hermann [310, Chap. 8]. Many applications in physics are to be found in Kirsten and Williams [354].

Exercise 1.4.5 (Convergence of Epstein's Zeta Function).

(a) Use Theorem 1.3.5 of Sect. 1.3 to show that Epstein's zeta function does indeed converge absolutely and uniformly on compact subsets of the half-plane $\operatorname{Re} s > n/2$.

(b) Develop an integral test in several variables to do the convergence proof in another way. We will return to this topic in Volume II [667].

Our next goal is to analytically continue Epstein's zeta function to the whole s-plane as a meromorphic function of s, having its unique pole at $s = n/2$. The method goes back to Riemann ([542, pp. 147 ff]) and this paper is translated and discussed in Edwards [149]. Riemann did not mention incomplete gamma functions, but he did, in fact, obtain this expansion when $n = 1$.

Theorem 1.4.1 (The Analytic Continuation of Epstein's Zeta Function). *Let $\Gamma(s,x)$ be the incomplete gamma function as in Exercise 1.4.3 and set*

$$G(s,x) \ = x^{-s}\Gamma(s,x) \ = \int_1^\infty y^{s-1}\exp(-xy)dy.$$

Then Epstein's zeta function can be analytically continued to all $s \in \mathbb{C}$ with its only pole a simple one at $s = n/2$ having residue $\frac{1}{2}\pi^{n/2}|Y|^{-1/2}\Gamma(n/2)^{-1}$. Here $|Y|$ denotes the determinant of Y.

The analytic continuation comes from **the incomplete gamma expansion:**

$$\begin{aligned}\Lambda(Y,s) &= \pi^{-s}\Gamma(s)Z(Y,s)\\ &= \frac{|Y|^{-1/2}}{2s-n} - \frac{1}{2s} + \frac{1}{2}\sum_{a\in\mathbb{Z}^n-0}\left(G(s,\pi Y[a]) + |Y|^{-1/2}G\left(\frac{n}{2}-s,\pi Y^{-1}[a]\right)\right).\end{aligned}$$

Thus Epstein's zeta function satisfies the **functional equation**

$$\Lambda(Y,s) = |Y|^{-1/2}\Lambda(Y^{-1},n/2-s).$$

Furthermore, Epstein's zeta function takes on the values

$$Z(Y,0) = -\frac{1}{2}, \quad Z(Y,-k) = 0,\ k = 1,2,3,\ldots.$$

Proof. This demonstration will be broken up into two exercises. □

Exercise 1.4.6 (The Transformation Formula of a Theta Function). For Y in $\mathcal{P}_n$, $t \in \mathbb{R}^+$, define

$$\theta(Y,t) = \sum_{a\in\mathbb{Z}^n}\exp(-\pi Y[a]t).$$

Show that the series converges and that

$$\theta(Y,t) = |Y|^{-1/2}t^{-n/2}\theta(Y^{-1},t^{-1}) \quad \text{with} \quad |Y| = \text{determinant of } Y.$$

Hint. *Use the same method as in Exercise 1.3.8 of Sect. 1.3. To do this you must write* $tY = B^2$*, with* $B \in \mathcal{P}_n$*. Then set* $f(x) = g(Bx)$*. The Fourier transforms of* f *and* g *are related by*

$$\hat{f}(x) = |B|^{-1}\hat{g}({}^tB^{-1}x).$$

Note. The series for theta converges quickly for large t. When t is near 0, the transformation formula allows one to replace t by t^{-1} and obtain a quickly convergent series again. We used this fact already in Theorem 1.3.5 of Sect. 1.3. It will be used over and over again in these notes.

Exercise 1.4.7 (Riemann's Trick).

(a) Show that for $\operatorname{Re} s > n/2$:

$$\Lambda(Y,s) = \frac{1}{2}\int_0^\infty t^{s-1}\left(\theta(Y,t) - 1\right)dt.$$

You can justify the interchange of summation and integration by Fubini's theorem (see Lang [387, p. 295]).

(b) Obtain the analytic continuation of Epstein's zeta function by writing

$$\int_0^\infty = \int_0^1 + \int_1^\infty .$$

Then replace t by $1/t$ in the first integral and use Exercise 1.4.6. This should lead you to the formula

$$\begin{aligned}\Lambda(Y,s) &= \int_1^\infty t^{s-1}w(Y,t)dt + |Y|^{-\frac{1}{2}}\int_1^\infty t^{n/2-s-1}w(Y^{-1},t)dt \\ &\quad + \frac{1}{2}\int_1^\infty \left(|Y|^{-\frac{1}{2}}t^{n/2} - 1\right)t^{-s-1}dt,\end{aligned}$$

where $w(Y,t) = \frac{1}{2}(\theta(Y,t) - 1)$. Then interchange sum and integral in the first two integrals and evaluate the last integral.

(c) Show that the incomplete gamma expansion of $Z(Y,s)$ converges exponentially faster than the original Dirichlet series defining $Z(Y,s)$.

Hint. *The incomplete gamma function has the asymptotic expansion*

$$G(s,x) \sim x^{-1}e^{-x}\left(1 + \frac{s-1}{x} + \frac{(s-1)(s-2)}{x^2} + \cdots\right), \quad \text{as } x \to \infty.$$

The proof of Theorem 1.4.1 is complete once you have done the preceding exercises. Much use has been made of such incomplete gamma expansions in number theory (see Montgomery and Vaughan [475] and Lavrik [398]). Our next goal is to apply the result to study the Riemann zeta function in the interval $(0,1)$. We do this in the next set of exercises, which go back to Riemann again (see Edwards [149, pp. 16–18]). Selberg generalized these exercises to the symmetric space $\mathcal{P}_n$ in order to obtain the analytic continuation of generalizations of Epstein's zeta function known as Eisenstein series. We will consider this in Volume II [667].

Exercise 1.4.8 (Multiplication Invariant Differential Operators on the Positive Reals).

(a) Let L be a differential operator acting on functions $f:\mathbb{R}^+\to\mathbb{C}$. Suppose $u:\mathbb{R}^+\to\mathbb{R}^+$ is differentiable with a differentiable inverse. Define L^u by $L^u f = L(f\circ u)\circ u^{-1}$. Let $a\in\mathbb{R}$ and $D_a = y^a(d/dy)y^{1-a}$. Then define $c(y)=cy$ for $c\in\mathbb{R}^+$. Show that $D_a^c = D_a$; i.e., D_a is multiplication invariant.

(b) Suppose that L is a multiplication invariant differential operator on $\mathbb{R}^+$. Define its formal adjoint L^* by

$$\int_{\mathbb{R}^+}(Lf)(y)\overline{g(y)}y^{-1}dy = \int_{\mathbb{R}^+} f(y)\overline{L^*g(y)}y^{-1}dy.$$

Note. dy/y is the multiplication invariant or **Haar measure** on the group of positive real numbers under multiplication. Show that if $u(y)=1/y$, then $L^u = \overline{L^*}$.

(c) Use integration by parts to prove that $D_a^u = -D_{2-a}$.

Exercise 1.4.9 (Transformation Formula of a Differentiated Theta Function). Set $L = D_{1/2}D_1$, using the notation of Exercise 1.4.8. Show that if $\theta(y,t)$ is the theta function of Exercise 1.4.6, when $n=1$, and $t,\ y\in\mathbb{R}^+$, then

$$L_t\theta(y,t) = y^{-\frac{1}{2}}t^{-\frac{1}{2}}L_t\theta(y^{-1},t^{-1}).$$

Exercise 1.4.10 (Power Series for ζ *(2s)*around $s=\frac{1}{4}$). If $L = D_{1/2}D_1$, then $L^* = D_1D_{3/2}$ and $L^*t^s = s(s-\frac{1}{2})t^s$. Then replace θ by $L_t\theta$ in the proof of Theorem 1.4.1 to show that

$$\begin{aligned}&2s(s-\tfrac{1}{2})\pi^{-s}\Gamma(s)\zeta(2s)\\ &\quad = \pi\textstyle\sum_{n\geq 1} n^2\int_{t=1}^{\infty}\left(t^s + t^{\frac{1}{2}-s}\right)\left(2\pi t n^2 - 3\right)\exp(-\pi t n^2)dt.\end{aligned}$$

Then show that we can expand the right-hand side as a power series in even powers of $s-\frac{1}{4}$ with positive coefficients. Deduce that $\zeta(s)<0$ for $0\leq s<1$.

The **Riemann hypothesis** says that $\zeta(s)$ has its only zeros at

$$s = -2,-4,-6,\ldots \quad \text{(the trivial zeros)}$$

and at points on the line $\operatorname{Re} s = \frac{1}{2}$, a line that is fixed by the functional equation of the zeta function. The hypothesis had been checked by computer for the first many million zeros by 1979 (see Brent [65] and Edwards [149]). By 2004 X.

Gourdon and P. Demichel had checked the Riemann hypothesis up to the 10^{13}th zero (see Ed Pegg, Jr.'s article called "Ten Trillion Zeta Zeros" at the website www.maa.org/editorial/mathgames). You win a million dollars if you have a proof of the Riemann hypothesis. Don't send the proof to me, please. See the Clay Math. Institute website (www.claymath.org). However, the proof or disproof of this hypothesis has so far skillfully eluded its many pursuers. In Volume II [667] we will discuss the evidence for the Riemann hypothesis coming from a study of spacings of the zeros (see also the paper on Odlyzko's website (www.dtc.umn.edu/~odlyzko/doc/zeta.htm).

Epstein's zeta function behaves very differently from Riemann's in some ways. For example, there are many matrices $Y \in \mathcal{P}_2$ such that $Z(Y,s)$ vanishes for some s with $\operatorname{Re} s > 1$ (see Davenport and Heilbronn [121]). And the following Exercise shows that the behavior of $Z(Y,s)$ for s in the interval $(0,1)$ is very different from that of Riemann's zeta function when Y has a small **minimum**

$$m_Y = \min\left\{Y[a] = {}^t aYa \mid a \in \mathbb{Z}^2 - 0\right\}.$$

Exercise 1.4.11. Use the incomplete gamma expansion of $Z(Y,s)$, $Y \in \mathcal{P}_2$, in Theorem 1.4.1 to graph the function $\Lambda(Y,s)$ for x in $(0,1)$ and

$$Y = \begin{pmatrix} t & 0 \\ 0 & 1/t \end{pmatrix}, \; t = 1, .1, .01, .001.$$

Hint. *You can compute the incomplete gamma functions using* Mathematica *or Matlab or whatever is your favorite mathematical software. Compare with Fig. 1.13.*

Historical Note.

Back in the day (more specifically, the 1970s and 1980s) we had to write our own programs to compute incomplete gamma functions and K-Bessel functions and the various special functions that came up in our work. My late ex-husband Riho Terras wrote the following ALGOL procedure to compute the incomplete gamma functions (see [673, 674]).

1. ALGOL PROCEDURE TO COMPUTE INCOMPLETE GAMMA FUNCTION G(S,X)

```
100 REAL PROCEDURE G(S,X);
200 COMMENT THIS FINDS INCOMPLETE GAMMA G(S,X) FOR 0<S<1
    AND X>0;
300 VALUE S,X; REAL S,X;
400 IF X GEQ 0.5 THEN BEGIN REAL W; INTEGER K;
500 FOR K:= ENTIER (8+50/X) STEP -1 UNTIL 1 DO
600 W:=X+W*(K-S)/(W+K);
700 G:= EXP(-X)/W;
800 END ELSE BEGIN
900 FOR K:=ENTIER(17+X+X) STEP -1 UNTIL 0 DO
```

```
1000 W:=(1+X*W)/(S+K);
1100 G:=X**(-S)*GAMMA(S)- EXP(-X)*W;
1200 END;
1300 END OF PROCEDURE G;
```

The procedure G is based on two formulas. When $0 < s < 1$ and $x > 0.5$, it uses the **continued fraction** for $G(s,x)$:

$$\begin{aligned} G(s,x) &= e^{-x}\left(\cfrac{1}{x+\cfrac{1-\mathrm{s}}{1+\cfrac{1}{x+\cfrac{2-\mathrm{s}}{1+\cfrac{2}{x+\dots}}}}}\right) \\ &\equiv e^{-x}\left(\frac{1}{x+}\,\frac{1-s}{1+}\,\frac{1}{x+}\,\frac{2-s}{1+}\,\frac{2}{x+}\cdots\right) \quad \text{for} \quad \mathrm{Re}\, s > 0. \end{aligned} \tag{1.22}$$

This is due to Legendre. See Exercise 1.4.12 for hints on a derivation. You can find a discussion of continued fractions in Henrici [289, Vol. II, Chap. 12, especially p. 629]. The ALGOL procedure quoted above views the continued fraction (1.22) as a composition of fractional linear transformations:

$$T_k(w) = x + w(k-s)/(w+k).$$

That is, the continued fraction shou1d be viewed as

$$T_1 T_2 \cdots T_k(w) = (e^x G(s,x))^{-1}, \tag{1.23}$$

for the correct choice of w. In the ALGOL procedure, we replace w by zero. This is analogous to using the first k terms of a series. The recursions sending w to $T_k(w)$ are error-correcting when $0 < s < 1$ and $x > 0$. The value of k used to compute $G(s,x)$ by truncating the continued fraction via (1.23) is $k = [8 + 50/x]$, where $[x]$ is the greatest integer $\leq x$, also known as the **floor** of x or $\lfloor x \rfloor$. Clearly this is very large if x is very near 0. In fact, the continued fraction is also bad when $s \geq 1$, because then the T_k start to magnify rather than correct errors. See R. Terras [673,674] and Henrici [289] for discussions of error-correcting recursions.

When $0 < s < 1$ and $0 < x < 0.5$, the ALGOL procedure above uses a power series:

$$\begin{aligned} &G(s,x) = x^{-s}\Gamma(s,x) = x^{-s}\Gamma(s) - x^{-s}\gamma(s,x), \\ &\text{when} \quad \gamma(s,x) = \int_0^x t^{s-1} e^{-t}\, dt = x^s e^{-x} \sum_{n \geq 0} \frac{x^n}{(s)_{n+1}}, \\ &(s)_0 = 1, \text{ and } (s)_n = \Gamma(s+n)/\Gamma(s) = s(s+1)\cdots(s+n-1). \end{aligned} \tag{1.24}$$

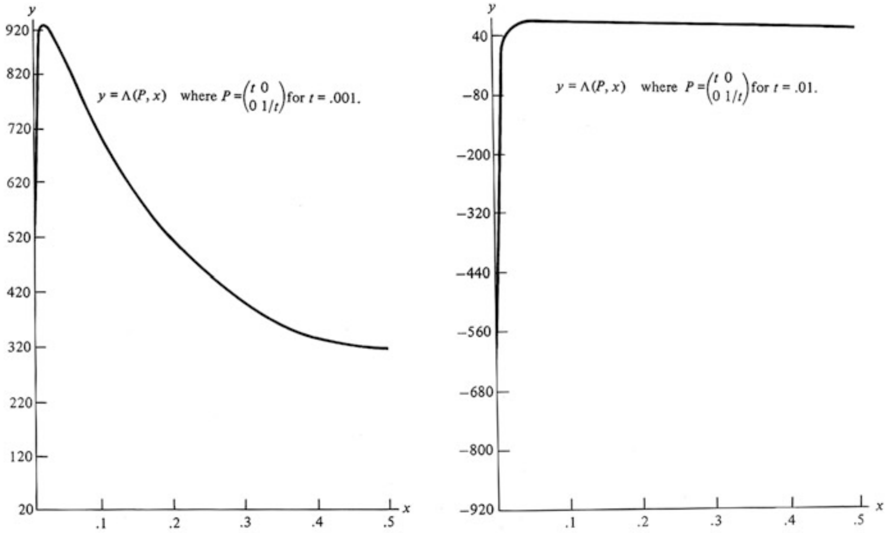


Fig. 1.13 Graphs of the product of the Epstein zeta function with its gamma factors in the interval $(0,1)$ from Purdy et al. [523]. Note that the x-axis does not cross the y-axis at $y = 0$. The figures show that the Epstein zeta functions graphed here vanish twice in the interval $(0,1)$. Thus the Epstein zeta function does not behave like Riemann's zeta function in the interval $(0,1)$

Exercise 1.4.12 (Properties of the Incomplete Gamma Function).

(a) Show that if $\varphi(s,x) = e^x G(s,x)$, then $\varphi(s+1,x) = 1 + s\varphi(s,x)/x$.
(b) Prove formula (1.22) above. This can be done using part (a) or via recursions for the confluent hypergeometric function.
(c) Prove formula (1.24) above.
(d) Use part (a) to prove the asymptotic expansion for $G(s,x)$ in the hint for Exercise 1.4.7.

Your answer to Exercise 1.4.11 should resemble the graphs in Fig. 1.13, which were produced in the 1970s by a Burroughs 6700 computer at the University of California (San Diego) using the ALGOL program for incomplete gamma functions above. See Purdy et al. [523].

Exercise 1.4.13 (Epstein Zeta Functions for Positive Definite Matrices with Small Minima over the Integer Lattice). Let $0 < u < 1$. Suppose Y is in $\mathcal{P}_n$ with $\det Y = |Y| = 1$. Let

$$m_Y = \min\left\{ Y[a] = {}^t aYa \mid a \in \mathbb{Z}^n - 0 \right\}.$$

Suppose that $m_Y \leq nu/2\pi e$ or $m_{Y^{-1}} \leq n(1-u)/2\pi e$. Show that if n is sufficiently large (dependent on u), $Z(Y, nu/2) > 0$. Conclude that $Z(Y,s)$ vanishes in $(nu/2, n/2)$.

Hint. *Use the fact that* $G(s,a) > a^{-s}\Gamma(s) - 1/s$. *Or see A. Terras [664].*

The size of m_Y in Exercise 1.4.12 is connected with part of Hilbert's 18th problem (see Milnor [468]) and we shall discuss it in more detail in Volume II [667]. It should be noted here that there are forms Y in $\mathcal{P}_n$ with $|Y| = 1$ and m_Y as small as you like if $n > 1$. Moreover, there are forms with $m_Y > n/(2\pi e)$, if n is large, as we shall prove in Volume II. In fact, we shall show that given s in $(0,\ n/2)$, there exist Y in $\mathcal{P}_n$ with $Z(Y,s)$ positive, negative, or zero. Related work is to be found in the paper of Sarnak and Strömbergsson [561].

Next we would like to consider some applications of Theorem 1.4.1 in algebraic number theory and crystallography. We shall discuss algebraic number theory first.

1.4.3 Algebraic Number Theory

Here we give a brief survey of the essentials of algebraic number theory. More details can be found in Hecke [260], Lang [388], Manin and Panchishkin [452], Narkiewicz [491], Samuel [554], Stark [623] and Stein [637] for example. There is a list of links to online number theory notes and teaching materials at

http://www.numbertheory.org/ntw/lecture_notes.html.

There are many computer programs for number theory such as PARI and SAGE. Mathematica will do some algebraic number theory computations. See H. Cohen [95] for more information on computational algebraic number theory.

Algebraic number theory originated with the work of Kummer, Dedekind, and others in the 1800s. They wanted to prove Fermat's conjecture that for $n = 3,4,5,\ldots$, the equation $x^n + y^n = z^n$ has no integer solutions x,y,z with $xyz \neq 0$. It was found that if one knows that the ring of cyclotomic integers $\mathfrak{O}_n = \mathbb{Z}[\exp(2\pi i/n)]$ has unique factorization into primes, then Fermat's conjecture follows. Unfortunately $\mathfrak{O}_n$ is only a unique factorization domain (UFD) for 29 of these rings (see Masley and Montgomery [455]). The Fermat conjecture was proved in 1995 by A. Wiles with the help of R. Taylor (see Wiles [740] as well as Taylor and Wiles [661]). We will have a little more to say about this in Chap. 3. See also Diamond and Shurman [133]. A popular introduction is to be found in Mozzochi's book [480].

Another question, which might appear initially to have nothing to do with cyclotomic integers, is deciding whether an ordinary integer is prime. This has become interesting to the CIA, NSA and all of us who carry out our business online because the first form of public key cryptography involves two huge primes, p and q, plus Fermat's little theorem (see Simmons [602] or Terras [668]). This method is called RSA cryptography, and RSA is now a company. There are attacks on RSA cryptosystems. The primes p and q must be very large The RSA website (www.rsa.com) has a discussion. The suggested size of the modulus $n = pq$ is 1024 bits (meaning that you must express the number pq base 2 and see something like $2^{1025} - 1$ zeros and ones). Thus the number of decimal digits is something like 308.

You would not like to see that number written down here. The RSA website says such a key is OK for corporate uses but not for extremely valuable keys.

Algorithms for distinguishing prime numbers from composite numbers are discussed in H. Cohen [95]. One such algorithm involves algebraic number theory in cyclotomic fields (see Adleman, Pomerance, and Rumely [2]). Algorithms for factoring are also discussed by Cohen.

Thus we are led to the study of **algebraic number fields** $K = \mathbb{Q}(\omega)$, ω **algebraic**; i.e., ω is a solution of an equation

$$f(\omega) = \omega^n + a_{n-1}\omega^{n-1} + \cdots + a_0 = 0,\ a_i \in \mathbb{Q}.$$

Assuming that the polynomial $f(x)$ is irreducible, we say that n is the **degree** of K over $\mathbb{Q}$. We shall write

$$n = [K : \mathbb{Q}].$$

Let $\omega^{(j)} \in \mathbb{C}$, $j = 1, 2, \ldots, n$, denote the roots of the irreducible polynomial $f(x)$ with $\omega^{(j)} \in \mathbb{R}$, for $j = 1, 2, \ldots, r_1$ and $\omega^{(j)} \notin \mathbb{R}$, for $j = r_1 + 1, \ldots, n$. Furthermore, we shall order the roots so that the non-real roots come in pairs of complex conjugates as follows:

$$\omega^{(j)} = \overline{\omega^{(j+r_2)}},\ j = r_1 + 1, \ldots, r_1 + r_2.$$

We always use an overbar to denote complex conjugate (and sometimes equivalence class or coset in a group). The map from ω to $\omega^{(j)}$ extends to a field isomorphism $x \to x^{(j)}$ of K into $\mathbb{C}$, fixing points of $\mathbb{Q}$. We shall call these isomorphisms the **conjugations** of K. The images of K in $\mathbb{C}$ are called conjugate fields to K. A **totally real number field** is such that all conjugate fields are subfields of $\mathbb{R}$, i.e., $r_2 = 0$ and $r_1 = n$.

The **trace** of $\alpha \in K$ is defined to be

$$\mathrm{Tr}_{K/\mathbb{Q}}\alpha = \mathrm{Tr}\alpha = \sum_{i=1}^{n} \alpha^{(i)}.$$

The **norm** of $\alpha \in K$ is defined to be

$$\mathrm{N}_{K/\mathbb{Q}}\alpha = \mathrm{N}\alpha = \prod_{i=1}^{n} \alpha^{(i)}.$$

One has the properties:

$$\mathrm{Tr}_{K/\mathbb{Q}}\alpha,\ \mathrm{N}_{K/\mathbb{Q}}\alpha \in \mathbb{Q},\ \ \mathrm{Tr}(\alpha+\beta) = \mathrm{Tr}(\alpha) + \mathrm{Tr}(\beta),\ \ \mathrm{N}(\alpha\beta) = \mathrm{N}\alpha\mathrm{N}\beta. \tag{1.25}$$

Problems involving the ordinary ring of integers $\mathbb{Z}$ in the rational number field $\mathbb{Q}$ lead directly to the study of the ring $\mathfrak{O}_K$ of **algebraic integers** α in the number field K; i.e., solutions of equations

$$\alpha^n + a_{n-1}\alpha^{n-1} + \cdots + a_0 = 0, \quad \text{with} \quad a_i \text{ in } \mathbb{Z}, \text{ all } i.$$

Number fields generalize ordinary arithmetic since $\mathfrak{O}_{\mathbb{Q}} = \mathbb{Z}$. And the arithmetic of these generalized integers is quite fascinating. It turns out that, just as in physics, analysis is needed to study these seemingly purely algebraic objects. A direct way to obtain algebraic information is to study the Dedekind zeta function of K and other related L-functions.

Since $\mathfrak{O}_K$ is not, in general, a UFD, one has to replace integers by ideals. An **ideal** $\mathfrak{A}$ in $\mathfrak{O}_K$ is an abelian group under addition such that $\mathfrak{O}_K\mathfrak{A} \subset \mathfrak{A}$ and $\mathfrak{A} \neq \{0\}$. A **principal ideal** has the form $\alpha\mathfrak{O}_K = (\alpha)$, for some α in $\mathfrak{O}_K$. The **product of two ideals** $\mathfrak{A}$ and $\mathfrak{B}$ is the ideal defined by

$$\mathfrak{A}\mathfrak{B} = \left\{ \sum_{i=1}^{k} a_i b_i \;\middle|\; a_i \in \mathfrak{A}, b_i \in \mathfrak{B} \right\}.$$

The sum **of two ideals** $\mathfrak{A}$ and $\mathfrak{B}$ is the ideal

$$\mathfrak{A} + \mathfrak{B} = \{ a + b \mid a \in \mathfrak{A}, b \in \mathfrak{B} \}.$$

The arithmetic of ideals works very much like the arithmetic of ordinary integers. One can show that every ideal can be factored uniquely (up to order) into a product of prime ideals.

Using elementary divisor theory (i.e., Gaussian elimination over $\mathbb{Z}$, which is the theory lying behind the fundamental theorem of abelian groups), one can show that every ideal $\mathfrak{A}$ has a **$\mathbb{Z}$-basis** or **integral basis**; i.e.,

$$\mathfrak{A} = \mathbb{Z}\alpha_1 \oplus \mathbb{Z}\alpha_2 \oplus \cdots \oplus \mathbb{Z}\alpha_n, \ \text{ for some } \ \alpha_1, \alpha_2, \ldots, \alpha_n \in \mathfrak{A}.$$

Thus we can view $\mathfrak{A}$ as an analogue of a crystal lattice. Moreover, one finds that the quotient $\mathfrak{O}_K/\mathfrak{A}$ is a finite ring. We will say that the **norm** of the ideal $\mathfrak{A}$ is

$$\text{Norm}\,\mathfrak{A} = \text{N}\mathfrak{A} = \#(\mathfrak{O}_K/\mathfrak{A}).$$

One can show that for a principal ideal (α), $N(\alpha) = |N\alpha|$, where the norm on the right is the product of conjugates of α.

If we look at the ideal $\mathfrak{O}_K = \mathbb{Z}\alpha_1 \oplus \cdots \oplus \mathbb{Z}\alpha_n$, then we can define the **discriminant** as

$$d_K = \det\left(\alpha_i^{(j)}\right)^2_{1 \le i,j \le n}.$$

To find a $\mathbb{Z}$-basis for $\mathfrak{O}_K$, one would want to use a computer with a number theory package such as SAGE. We give a short table of examples at the end of our brief review of number theory. The divisors of d_K are the primes p in $\mathbb{Z}$ such that the ideal

$p\mathfrak{O}_K$ factors into prime ideals in K with squares of prime ideals appearing. This is Dedekind's discriminant theorem (see Samuel [554, p. 75]).

As we said, one wants to know if $\mathfrak{O}_K$ is a UFD for many applications. To measure how far $\mathfrak{O}_K$ deviates from unique factorization, Dedekind defined the **ideal class group** I_K of ideals modulo the equivalence relation

$$\mathfrak{A} \sim \mathfrak{B} \Leftrightarrow \gamma\mathfrak{A} = \mathfrak{B}, \quad \text{for some} \quad \gamma \neq 0 \text{ in } K.$$

Multiplication of ideals induces a group operation on I_K which makes I_k a finite group whose order h_K is called the **class number** of K. The ring $\mathfrak{O}_K$ is a UFD if and only if $h_K = 1$. We have only considered ideals which are subsets of $\mathfrak{O}_K$. But it is also useful to enlarge this definition. A **fractional ideal** $\mathfrak{A}$ is a subset of K forming an abelian group under addition such that $\mathfrak{O}_K\mathfrak{A} \subset \mathfrak{A}$ and $\alpha\mathfrak{A} \subset \mathfrak{O}_K$ for some $\alpha \neq 0$ in $\mathfrak{O}_K$. The arithmetic of fractional ideals is very like that of the rational numbers $\mathbb{Q}$.

It is usually fairly hard to tell whether the class number of a given field is one. For imaginary quadratic fields $\mathbb{Q}(\sqrt{d})$ the answer was obtained during the 1960s (see Stark [625]). The imaginary quadratic fields $\mathbb{Q}(\sqrt{d})$ with class number one have discriminant $d = -3,\ -4,\ -7,\ -8,\ -11,\ -19,\ -43,\ -67,\ -163$. Proofs of this statement involve analysis, using either Epstein zeta functions or modular forms. The theorem was suspected for a long time because it is easy to produce big tables of class numbers for those fields. We will show how to do this in Chap. 3. See Watkins [719] for the solution of the problem of finding all imaginary quadratic fields of class number ≤ 100.

For real quadratic fields, the situation is very different. Tables of class numbers lead one to believe that there should be an infinite number of real quadratic fields with class number one, but no one can prove it (see Stein [637, p. 81], for a table of real quadratic fields of class number 1, or Williams and Broere [742]). Masley and Montgomery (see [455]) found all the cyclotomic fields with class number one. Previously Siegel had reduced the problem of determining the prime cyclotomic fields of class number one to a finite problem (see Siegel [600, Vol. III, p. 442]).

Note that there is a problem computing with ideals, since they are infinite sets of numbers. Thus one often wants to go up to a field $E \supset K$ such that the ideals of K become principal in E. Here E may not have class number one but, in any case, one can do the desired computation more easily in E. Hecke had to do this many times (see Hecke [258, p. 255ff], for example). Class field theory shows that E exists (see Cassels and Fröhlich [79, Exercise 3, pp. 355–357]).

There is a more surprising invariant of a number field waiting to be described—the **group of units** $U_K = \left\{u \in \mathfrak{O}_K | u^{-1} \in \mathfrak{O}_K\right\}$. This group is not very interesting if $K = \mathbb{Q}$, since it then consists of $+1$ and -1. It is also not very interesting for an imaginary quadratic field since if m is square-free and negative then $U_{\mathbb{Q}(\sqrt{m})} = \{+1,\ -1\}$ unless $m = -1$ or -3. If $m = -1$, you get the fourth roots of unity. If $m = -3$, you get the sixth roots of unity. However, the real quadratic fields contain infinite cyclic groups of units as well as $+1$ and -1. A generator of the infinite cyclic group of units is called a **fundamental unit** for the real quadratic field. You can find a fundamental unit for $\mathbb{Q}(\sqrt{d})$, when $d > 0$, using the continued fraction expansion for $\sqrt{d}$.

In general, Dirichlet's unit theorem says that the group of units has the form

$$U_K = \langle \varepsilon_1 \rangle \times \cdots \times \langle \varepsilon_r \rangle \times W,$$

where $\langle \varepsilon_j \rangle$ denotes the infinite cyclic group generated by ε_j, W is the group of roots of unity in K (a finite cyclic group), $r = r_1 + r_2 - 1$, with r_i as defined at the very beginning of the discussion of algebraic number fields. The ε_i are called **fundamental units** of K. The measure of the unit group is given to a certain extent by the **regulator** defined by

$$R_K = \det \left(\log \left| \varepsilon_i^{(j)} \right|^{e_j} \right)_{1 \le i,j \le r}$$

with

$$e_j = \begin{cases} 1, & j = 1, \ldots, r_1 \\ 2, & j = r_1 + 1, \ldots, r. \end{cases}$$

Since the norm of a unit is ± 1, it does not matter which embedding $K \to \mathbb{C}$ is left out in the definition of the regulator.

Minkowski's lemma in the geometry of numbers (see Volume II [667] or Samuel [554, p. 55]) gives the standard proof of the finiteness of the class number as well as Dirichlet's unit theorem. One can also deduce these things from the convergence of Dedekind's zeta function (see Theorem 1.4.2 and Siegel [600, Vol. II, pp. 93–94]).

Note that if $r = 1$, the regulator is the logarithm of a fundamental unit. This is a transcendental number (see Baker [21, p. 6]), as was proved by Lindemann in 1882. Thus one might expect the same of the regulator itself, but no one knows how to prove that.

The search for fundamental units is a problem that we can only begin to attack after we have considered the results of Volume II. The connection between algebraic number theory and that of the general linear group of $n \times n$ nonsingular real matrices is basic to much work on this subject (see, e.g., T. Shintani [593]).

Conjectures of H. Stark allow one to construct units of class fields of number fields from values of derivatives of L-functions. The units are not fundamental, in general, but the conjecture allows one to find the units explicitly in many interesting cases. This also relates to the 12th in Hilbert's famous list of mathematical problems for the turn of the last millennium (see Stark [626, 627] or C. Popescu, K. Rubin and Alice Silverberg [519]).

One of the major problems of algebraic number theory is describing the growth of the product of the class number times the regulator as the discriminant goes to infinity. We shall see why one looks at the class number times the regulator after Theorem 1.4.2. In fact, it is hard to separate h and R with the available tools. The **Brauer-Siegel theorem** says that under certain hypotheses on the sequence of fields (e.g., fixed degree over $\mathbb{Q}$), one has

$$\log(hR) \sim \log |d|^{1/2} \quad \text{as } |d| \to \infty.$$

Table 1.4 Examples of number fields

K	$\mathfrak{O}_K$	d_K	U_K	h_K
$\mathbb{Q}(\sqrt{m})$ $m<0$ square-free	$\mathbb{Z}[\sqrt{m}]$ if $m \equiv$ $2,3(\mathrm{mod}\,4)$	m or $4m$	$\{1,-1\}$ $m \neq -1$ or -3	$=1$ for only 9 fields
$\mathbb{Q}(\sqrt{m})$ $m>0$ square-free	$\mathbb{Z}[\sqrt{m}]$, if $m \equiv$ $2,3(\mathrm{mod}\,4)$	m or $4m$	$\{1,-1\} \times \langle \varepsilon \rangle$ find ε by continued fractions	conjectured =1 for infinitely many fields
$\mathbb{Q}(\sqrt[3]{m})$ m cube-free	$\mathbb{Z}[\sqrt[3]{m}]$ sometimes	$3^a m$ sometimes	$\{1,-1\} \times \langle \varepsilon \rangle$ find ε by Voronoi's algorithm	conjectured =1 for infinitely many fields
$\mathbb{Q}\left(e^{2\pi i/m}\right)$	$\mathbb{Z}\left[e^{2\pi i/m}\right]$	$=\pm p^{p-2}$ if $m=p$ $=$ prime	$r=\frac{\varphi(m)}{2}-1$ $\varphi=$ Euler function	=1 for only 29 distinct fields

Thus, in particular, for imaginary quadratic fields (where $R=1$), there can be only a finite number of fields with a given class number. Siegel did the quadratic version of the theorem. Brauer did the more general result using the theory of representations of finite groups and Artin L-functions. The whole proof would be trivial if one knew that the Dedekind zeta function (defined below) has no real zeros near 1 (see Lang [388]). Unfortunately the behavior of the Dedekind zeta function near $s=1$ is still a mystery. The possible zeros near 1 are called "**Siegel zeros**." References for them include Purdy et al. [523], Stark [624], and Siegel [598]. See Odlyzko [503] for a survey of work on bounds on discriminants and estimates of class numbers and regulators.

Table 1.4 gives some examples of number fields and their invariants. References for this table are Narkiewicz [491], Stark [624], Williams and Broere [742], Barrucand, Williams, and Baniuk [27], and Masley and Montgomery [455]. See Cohen [95] or Stein [637] for more information on computational algebraic number theory.

The **Dedekind zeta function** of a number field K is defined to be

$$\zeta_K(s) = \sum_{\text{ideals } \mathfrak{A} \subset \mathfrak{O}_K} N\mathfrak{A}^{-s} = \prod_{\text{prime ideals } \mathfrak{P}} (1 - N\mathfrak{P}^{-s})^{-1}$$

for $\operatorname{Re} s > 1$. Unique factorization of ideals into prime ideals gives the Euler product above just as in Exercise 1.4.4 for the case $K=\mathbb{Q}$. For $\zeta_{\mathbb{Q}}(s)$ is nothing other than Riemann's zeta function.

The convergence of Dedekind's zeta function can be proved algebraically using a fact from algebraic number theory. One needs to know that if p is a prime in $\mathbb{Z}$, then

$$p\mathfrak{O}_K = \mathfrak{P}_1^{e_1} \cdots \mathfrak{P}_g^{e_g}, \quad \text{with prime ideals } \mathfrak{P}_i \text{ in } \mathfrak{O}_K,$$

such that $N\mathfrak{P}_i = p^{f_i}$. Now $p^n = N(p) = p^{\Sigma f_i e_i}$ and it follows that

$$n = \sum_{i=1}^{g} f_i e_i.$$

It is also clear that if $\mathfrak{P}$ is a prime ideal in K then $\mathfrak{P} \cap \mathbb{Z} = (p)$, where p is a prime number in $\mathbb{Z}$. This shows (using the Euler product for ζ_K) that for $s > 1$

$$\log \zeta_K(s) = \sum_{\mathfrak{P}} \sum_{k \geq 1} k^{-1} N\mathfrak{P}^{-ks} \leq n \sum_{p} \sum_{k \geq 1} k^{-1} p^{-ks} = n \log \zeta(s).$$

So we have

$$\zeta_K(s) \leq (\zeta(s))^n \quad \text{when} \quad s > 1. \tag{1.26}$$

In 1916, Hecke figured out how to continue $\zeta_K(s)$ to the whole complex s-plane as a meromorphic function with a simple pole at $s = 1$ and residue

$$\operatorname*{Res}_{s=1} \zeta_K(s) = 2^{r_1} (2\pi)^{r_2} h R w^{-1} |d|^{-\frac{1}{2}},$$

where r_1, r_2 are the numbers of real conjugates and pairs of complex conjugates of K,respectively, h is the class number, R the regulator, d the discriminant, and w the number of roots of unity in K. This clearly relates class number-regulator problems to the Dedekind zeta function. Moreover, Hecke's proof can easily be seen to connect these problems with the behavior of the Dedekind zeta function in the interval $(0,1)$.

Landau used Hecke's result (plus Hadamard factorization and Cauchy's integral theorem) to prove the **prime ideal theorem**:

$$\#\{\text{prime ideals } \mathfrak{P} \text{ in } K \mid N\mathfrak{P} \leq x\} \sim x/\log x \text{ as } x \to \infty.$$

This generalizes the ordinary prime number theorem when $K = \mathbb{Q}$. There is a proof in Goldstein [223]. See also Lagarias and Odlyzko [383].

We want to relate $\zeta_K(s)$ and Epstein's zeta function $Z(Y,s)$ using a method devised by Hecke [258, pp. 198–207]. This method embodies the essence of the relation between the general linear group and algebraic number fields. It will tell us all the properties of the Dedekind zeta function.

Theorem 1.4.2 (Hecke's Relation Between Dedekind and Epstein Zeta Functions). *Let $\varepsilon_1, \ldots, \varepsilon_r$ be a system of fundamental units for the algebraic number field K. For $x \in \mathbb{R}^r$, set*

$$\tau_j = \prod_{i=1}^{r} |\varepsilon_i^{(j)}|^{2x_i}, \quad j = 1, 2, \ldots, n.$$

Let C be an ideal class in the ideal class group I_K of K. For any ideal $\mathfrak{A}$ in C, let $\mathfrak{A} = \mathbb{Z}\omega_1 \oplus \mathbb{Z}\omega_2 \oplus \cdots \oplus \mathbb{Z}\omega_n$. Then define the matrix $Y(C,x) \in \mathcal{P}_n$ by

$$Y(C,x) = \begin{pmatrix} \tau_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \tau_n \end{pmatrix} \left\{ \begin{pmatrix} \omega_1^{(1)} & \cdots & \omega_n^{(1)} \\ \vdots & \ddots & \vdots \\ \omega_1^{(n)} & \cdots & \omega_n^{(n)} \end{pmatrix} \right\},$$

using the notation $Y\{A\} = {}^t\bar{A}YA$. If $\zeta_K(s)$ is Dedekind's zeta function, set

$$\Lambda_K(s) = (2^{-r_2}|d_K|^{1/2}\pi^{-n/2})^s \Gamma(s/2)^{r_1}\Gamma(s)^{r_2}\zeta_K(s).$$

Here the degree is $n = [K:\mathbb{Q}]$, and r_1 is the number of real conjugate fields of K, r_2 the number of pairs of complex conjugate fields of K. Let $Z(Y,s)$ be Epstein's zeta function and

$$\Lambda(Y,s) = \pi^{-s}\Gamma(s)Z(Y,s).$$

We also use the notation $Y^0 = |Y|^{-1/n}Y$ for Y in $\mathcal{P}_n$. Clearly $|Y^0| = 1$. Then ***Hecke's integral formula*** *is*

$$\frac{w}{2^{r_1}nR}\Lambda_K(s) = \sum_{C \in I_K} \int_{x \in [0,1]^r} \Lambda\left(Y(C,x)^0, \frac{ns}{2}\right) dx.$$

Here w is the number of roots of unity in K and$R$ is the regulator of K.

Proof. **Step 1. The switch from a sum over ideals to a sum over elements of ideals.** Recall that I_K is the ideal class group. Let C be an ideal class in I_K. Choose an ideal $\mathfrak{A}$ in the inverse class C^{-1}. Then $\mathfrak{A}\mathfrak{B} = \alpha\mathfrak{O}_K = (\alpha)$ for all ideals $\mathfrak{B}$ in C. Thus

$$\zeta_K(s) = \sum_{C \in I_K} N\mathfrak{A}^s \sum_{\alpha \in (\mathfrak{A}-0)/U_K} N(\alpha)^{-s},$$

where U_K is the group of units in $\mathfrak{O}_K$ and the quotient $(\mathfrak{A}-0)/U_K$ means take representatives for the equivalence relation $\alpha \sim \beta$ if and only if $\alpha = \beta u$ for some u in U_K. Here we have used the multiplicativity of the norm and the fact that $(\beta) = (\alpha)$ if and only if $\beta = \alpha u$, for some u in U_K.

Step 2. Products of Γ-functions as products of Mellin transforms. The Mellin transform formula for the gamma function is

$$a^{-s}\Gamma(s) = \int_{\mathbb{R}^+} y^{s-1}\exp(-ay)dy, \quad \text{for } a > 0,\ Re\ s > 0. \tag{1.27}$$

Taking products of such integrals, we obtain

$$\left(2^{-r_2}|d_K|^{1/2}\pi^{-n/2}\right)^s \Gamma(s/2)^{r_1}\Gamma(s)^{r_2}\left(N\mathfrak{A}/N(\alpha)\right)^s$$
$$= \int_{y\in(\mathbb{R}^+)^{r_1+r_2}} Ny^{s/2}\exp\left(-\pi\left(|d_K|N\mathfrak{A}^2\right)^{-1/n}\ \mathrm{Tr}\left(\overline{\alpha}y\alpha\right)\right)\frac{dy}{y},$$

where

$$\mathrm{N}y = \prod_{j=1}^{r_1+r_2} y^{e_j}, \quad \mathrm{Tr}\, y = \sum_{j=1}^{r_1+r_2} e_j y_j, \ \frac{dy}{y} = \prod_{j=1}^{r_1+r_2}\frac{dy_j}{y_j},$$

and

$$e_j = \begin{cases} 1, \text{ for } j = 1,\ldots,r_1, \\ 2, \text{ for } j = r_1+1,\ldots,r_1+r_2. \end{cases}$$

The definitions are set up to make

$$Ny = \prod_{j=1}^{r_1+r_2} |\alpha^{(j)}|^{e_j} = \mathrm{N}(\alpha), \quad \text{if } y = \left(|\alpha^{(1)}|,\ldots,|\alpha^{(r_1+r_2)}|\right),$$

for example. When we write $\bar{\alpha}y\alpha$ we mean the element of $(\mathbb{R}^+)^{(r_1+r_2)}$ given by the vector with jth component $|\alpha^{(j)}|^2 y_j$.

Step 3. Hecke's change of variables. Hecke decided to send $(y_1, \ldots, y_{r_1+r_2})$ to $(u,x_1, \ldots,x_r)$ via the equations

$$y_j = u\tau_j, \quad j = 1,\ldots, r_1+r_2.$$

Here the τ_j are defined in the statement of the theorem as products of powers of conjugates of fundamental units. If the fundamental units did not exist, then we would be able to deduce a contradiction to the convergence of Dedekind's zeta function. Exercise 1.4.14 below is needed to perform the indicated substitution in the integral of step 2. Thus we obtain

$$\left(2^{-r_2}|d_K|^{1/2}\pi^{-n/2}\right)^s \Gamma(s/2)^{r_1}\Gamma(s)^{r_2}\left(N\mathfrak{A}/N(\alpha)\right)^s$$
$$= n2^{r_1-1}R\int_{x\in\mathbb{R}^r}\int_{u>0} u^{-1+ns/2}\exp\left(-\pi u\left(|d_K|N\mathfrak{A}^2\right)^{-1/n}\ \mathrm{Tr}(\overline{\alpha}\tau\alpha)\right)du\, dx.$$

Here $\tau = (\tau_1, \ldots, \tau_n)$.

Step 4. Perform the integral over u and switch the quotient modulo units to the domain of the integral over x. You obtain

$$\left(2^{-r_2}|d_K|^{1/2}\pi^{-n/2}\right)^s \Gamma(s/2)^{r_1}\Gamma(s)^{r_2}\zeta_K(s)$$
$$= n2^{r_1}Rw^{-1}\pi^{-ns/2}\Gamma(ns/2)|d_K|^{s/2}\sum_{C\in I_K}\int_{x\in[0,1]^r} Z\left(Y(C,x),ns/2\right)\, dx.$$

To see this, note that the Dirichlet unit theorem says that if $u \in U_K$ then $u = \kappa \varepsilon_1^{a_1} \cdots \varepsilon_r^{a_r}$ *for* some $a_i \in \mathbb{Z}$ where κ is a root of unity in K. This decomposition is unique. Thus

$$\left|(\alpha u)^{(j)}\right|^2 \tau_j = \left|\alpha^{(j)}\right|^2 \prod_{i=1}^{r} \left|\varepsilon_i^{(j)}\right|^{2(x_i+a_i)}.$$

This completes the proof, once you do the following pair of exercises.

□

Exercise 1.4.14. Show that the Jacobian of Hecke's change of variables in step 3 above is

$$\left|\frac{\partial y}{\partial(x,u)}\right| = y_1 \cdots y_{r+1} u^{-1} 2^{r_1 - 1} nR.$$

Exercise 1.4.15. Verify that $Y(C,x)$ is actually a positive definite symmetric matrix.

Hint. *Suppose that α in $\mathfrak{A}$ has the form $\alpha = \Sigma a_j \omega_j$ for $a = {}^t(a_1, \ldots, a_n)$ in $\mathbb{Z}^n$. Then $Y(C,x) = \mathrm{Tr}(\overline{\alpha}\tau\alpha)$. This gives a positive real number even if you plug in a from $\mathbb{R}^n$. To see that the matrix $Y(C,x)$ is real and symmetric, you need to look at the permutation*

$$\sigma = \begin{pmatrix} 1 & \cdots & r_1 & r_1+1 & \cdots & r_1+r_2 & r_1+r_2+1 & \cdots & n \\ 1 & \cdots & r_1 & r_1+r_2+1 & \cdots & n & r_1+1 & \cdots & r_1+r_2 \end{pmatrix}$$

induced by complex conjugation on the conjugates of K. Then $\tau_{\sigma(j)} = \tau_j$.

Note that the value of the Epstein zeta function in Hecke's integral formula is independent of the choice of ideal in C and integral basis of that ideal. It is also independent of the particular set of fundamental units that we choose. In particular, $|Y(C,x)| = \mathbb{N}\mathfrak{A}^2 |d_K|$, since the integer matrix relating a $\mathbb{Z}$-basis of $\mathfrak{A}$ and a $\mathbb{Z}$-basis of $\mathfrak{O}_K$ can be diagonalized by elementary divisor theory using elementary row and column operations. Changing the ideal in C which is used replaces the matrix $Y(C,x)$ by $Y(C,x)[A] = {}^tAY(C,x)A$, for some A in $GL(n,\mathbb{Q})$. Since we are taking determinants to be one in the integral formula, we are transforming by $GL(n,\mathbb{Z})$, which leaves the Epstein zeta function invariant. A similar argument applies to changing the $\mathbb{Z}$-basis of the ideal.

If the field K is imaginary quadratic, the formula in Theorem 1.4.2 is much simpler, since then there are no integrals. Then it is in fact a very old formula, expressing the connection between ideals in imaginary quadratic fields and binary quadratic forms. We will see in Chap. 3 that this leads to an algorithm for the computation of class numbers of imaginary quadratic fields. This explains how Gauss could make class number conjectures before class numbers were invented.

He was talking about class numbers of binary quadratic forms and conjectured that in the positive definite case there are only finitely many discriminants with a given class number.

Corollary 1.4.1 (Properties of the Dedekind Zeta Function).

(1) $\zeta_K(s)$ can be continued to a meromorphic function of s with a simple pole at $s=1$ having residue

$$\operatorname*{Re}_{s=1} s\ \zeta_K(s) = 2^{r_1}(2\pi)^{r_2} hRw^{-1}|d|^{-1/2}$$

and $\zeta_K(s)$ satisfies the ***functional equation***

$$\Lambda_K(s) = \left(2^{-r_2}|d|^{1/2}\pi^{-n/2}\right)^s \Gamma(s/2)^{r_1}\Gamma(s)^{r_2}\zeta_K(s) = \Lambda_K(1-s).$$

(2) The Dedekind zeta function has the ***incomplete gamma expansion****:*

$$\Lambda_K(s) = \frac{2^{r_1}hR}{ws(s-1)} + \frac{n2^{r_1-1}R}{w}\sum_{\substack{C\in I_K\\ a\in\mathbb{Z}^n-0}} \int\limits_{x\in[0,1]^r} \left(G\left(\frac{ns}{2}, Y(C,x)^0[a]\right)\right.$$
$$\left. + G\left(\frac{n(1-s)}{2}, Y(C,x)^0[a]\right)\right) dx.$$

Proof. The corollary follows from Theorems 1.4.1 and 1.4.2, using the fact that $Y(C,x)^{-1} = Y(C', -x)$, where C' denotes the ideal class containing the ideal $\mathfrak{A}'$ dual to the ideal $\mathfrak{A}$ in C that we had selected. If $\mathfrak{A}$ is an ideal the **dual ideal** $\mathfrak{A}'$ is defined to be

$$\mathfrak{A}' = \{\beta \in K | \mathrm{Tr}(\alpha\beta) \in \mathbb{Z} \text{ for all } \alpha \in \mathfrak{A}\}.$$

See Lang [388, pp. 57–58], where it is shown that $\mathfrak{A}'$ is a fractional ideal in general. You need to see that if ω' is a dual basis for the $\mathbb{Z}$-basis ω of the ideal i.e., if

$$\mathrm{Tr}\ (\omega_i\omega_j') = \delta_{ij} = \text{the Kronecker delta},$$

then $\left(\omega_i^{(j)}\right)^{-1} = {}^t\left(\omega_i'^{(j)}\right)$ and ω' is a $\mathbb{Z}$-basis for the dual ideal to that generated by ω. □

The most important dual ideal is that for $\mathfrak{O}_K$ itself. The **different** is defined to be $\mathfrak{d}_K = (\mathfrak{O}_K')^{-1}$. And $N\mathfrak{d}_K = |d_K|$. The concept of **dual ideal** is really the same as the concept of dual lattice or dual subgroup of a locally compact abelian group. Poisson summation will always relate the sum over a lattice and that over the dual lattice. In the special case under consideration here our lattice is an ideal in $\mathfrak{O}_K$. We can view this as a subset of $\mathbb{R}^n$ using the conjugations. More generally, a subset L of $\mathbb{R}^n$ is a **lattice** if L is a discrete subgroup of $\mathbb{R}^n$ such that $\mathbb{R}^n/L$ has finite volume. More information on Euclidean lattices can be found in Siegel [597]. The **dual lattice** L^* to a lattice L in $\mathbb{R}^m$ consists of the vectors u such that ${}^t uv$ is in $\mathbb{Z}$ for all v in L. For example, if $n=1$, $L=a\mathbb{Z}$, $a\neq 0$,and $L^*=\frac{1}{a}\mathbb{Z}$. See also Weil [726, Chap. 2]. When one replaces lattices by positive definite matrices, the dual lattice corresponds to the inverse matrix.

If the class number were infinite or if the r fundamental units did not exist, then one could show that $\zeta_K(s)$ is infinite, contradicting the bound that we found when Re $s>1$ in formula (1.26). This point of view is developed in Siegel [600, Vol. II, pp. 93–94], where it is noted that Dirichlet discovered his unit theorem "als er im Jahre 1846 in Rom der Ostermusik in der Sixtinischen Kapelle zuhörte."

We see from the corollary that $\zeta_K(s)$ is negative when s is less than, but sufficiently close to 1. It might be expected that the Dedekind zeta function would behave like Riemann's and thus its only possible zero in (0,1) would be at the point $\frac{1}{2}$, but no one has been able to prove this, even for quadratic fields (cf. Stark [624], Purdy et al. [523], Stark and Zagier [634]). In fact, one expects that any proof that works for the Riemann zeta function should be generalizable to the Dedekind zeta function. For example, Siegel says (see [600, Vol. II, p. 72]), "Sollte einmal ein Beweis für die Richtigkeit der Riemannschen Vermutung gelingen, so wird dieser vermutlich für alle, auch die von Hecke verallgemeinerten ζ-Funktionen gültig sein." However, although this statement appears true for the case of Landau's prime ideal theorem, it does not appear to work for real zeros of the Dedekind zeta function. For it is easy to show that the Riemann zeta function has no zeros in $(0,1)$ (cf. Exercise 1.4.10). But no one knows how to generalize this to Dedekind's zeta function (i.e., there may exist **"Siegel zeros"**).

Of course one still does not know whether the Riemann hypothesis holds for the Dedekind (or Riemann) zeta functions. Most computer work has been done for Riemann's zeta function, but see Lagarias and Odlyzko [383] and the website www.lmfdb.org, which is a database for L-functions, modular forms, and related objects.

We are interested in the sign of $\zeta_K(s)$ in $(0,1)$ thanks to the Brauer–Siegel theorem. For the incomplete gamma expansion in the corollary allows one to say that $\zeta_K(s)\leq 0$ implies

$$\frac{2^{r_1}w^{-1}hR}{s(1-s)} \geq \text{ a sum of integrals of incomplete gamma functions.}$$

The best way to analyze the terms on the right-side of this inequality seems to involve rewriting the terms as higher-dimensional incomplete gamma functions. If we can get good lower bounds on these integrals, then we get good lower bounds on the product of the class number and the regulator. Upper bounds come from (1.26), for example.

It is tempting to try to use Hecke's integral formula (Theorem 1.4.2) to deduce the behavior of $\zeta_K(s)$ for s in $(0,1)$ from the behavior of $Z(Y,s)$ for s in $(0,n/2)$.

But we know from Exercises 1.4.11 and 1.4.13 that the behavior of the Epstein zeta function is very different from the expected behavior of Dedekind's zeta function.

One can also interpret Hecke's integral formula (Theorem 1.4.2) as the computation of the 0th Fourier coefficient of the function $\Lambda_n(Y(C,x),ns/2)$, considered as a periodic function of x. Siegel computes all the Fourier coefficients for K a real quadratic field in [599]. One obtains Hecke L-functions with grossencharacters χ:

$$L(s,\chi) = \sum \chi(\mathfrak{A}) N\mathfrak{A}^{-s}.$$

For a real quadratic field of class number one and fundamental unit ε, an example of a Hecke grossencharacter is

$$\chi((\alpha)) = (\alpha/\alpha')^{ni/\log \varepsilon}, \quad \text{for } \alpha \text{ in } \mathfrak{O}_K.$$

Here α' means the conjugate of α in K. For more information on Hecke grossencharacters, see Hecke [258, pp. 215–234, 249–287].

If K is imaginary quadratic, there is no integral in Hecke's integral formula. Then the result analogous to that of Siegel gives a connection between nonanalytic Eisenstein series and Hecke L-functions with grossencharacters. An example of a grossencharacter χ for an imaginary quadratic field of class number one is:

$$\chi((\alpha)) = \alpha^e \overline{\alpha}^f, \quad \text{with } e+f \text{ even}.$$

In 1970 Damerell found $L(n,\chi)$, for large n in $\mathbb{Z}^+$, to be an algebraic number times a certain explicit transcendental number (see Weil [727]).

You might ask why one cares about values of L-functions at positive integers. The earliest result of this type was that of Euler, who showed that for Riemann's zeta function (see Exercise 3.5.7 of Sect. 3.5),

$$\zeta(2m) = (-1)^{m-1}\frac{(2\pi)^{2m}B_{2m}}{2(2m)!},\ m = 1,2,3,\ \ldots.$$

Here $B_n \in \mathbb{Q}$ is the nth **Bernoulli number** defined by

$$\frac{x}{e^x-1} = \sum_{n\geq 0} B_n \frac{x^n}{n!}.$$

Analogous results, using similar methods, were obtained by Shintani [593, 595] for the Dedekind zeta function of a totally real field. Interpretations involving p-adic interpolation of these values have been much discussed (see Coates [94] or Iwasawa [321]). There are also some very general conjectures of Lichtenbaum giving K-theoretic meaning to values of L-functions at negative integers (see Borel [51], Lichtenbaum [420], Serre [578]).

Values of L-functions are of interest also because of their connection with Hilbert's 12th problem which asks for an explicit construction of class fields E or abelian extensions of number fields K. Here "abelian" means the Galois group of E/K is commutative or abelian. When $K = \mathbb{Q}$, a theorem of Kronecker and Weber says that a finite extension E of $\mathbb{Q}$ with abelian Galois group must lie in a cyclotomic field $E \subset \mathbb{Q}(\exp(2\pi i/m))$. When the base field is imaginary quadratic rather than $\mathbb{Q}$, one adjoins values of modular functions (see Shimura [589], Borel and Chowla [53]). When K is arbitrary, Shimura has obtained results of this sort using algebraic geometry. Another point of view is expressed in Stark [626]. Stark's idea is that values of L-functions can be used to find units which generate class fields and express the reciprocity laws of class field theory. He has proved this in classical cases and has convincing computer examples in other cases (e.g., the base field real quadratic or cubic). See also Popescu et al. [502].

1.4.4 Crystallography

A three-dimensional lattice is the set of all linear combinations with integer coefficients of three linearly independent vectors v_1, v_2, v_3 in $\mathbb{R}^3$. If you choose the standard basis vectors of $\mathbb{R}^3$, your lattice is $\mathbb{Z}^3$. If you think of atoms as points with legs, then by connecting atoms you can build up a crystal lattice such as that of salt (NaCl) pictured along with various other lattices in Figs. 1.14 and 1.15 from Born and Huang [57, pp. 383–384].

A **crystallographic point group** G is a subgroup of the group $O(3)$ of rotations of 3-space which leave a lattice L invariant; i.e., for each v in L the vector gv is in L, for any g in G. Note that you can represent the elements of G by matrices with integer entries or by rotation matrices. Thus G is compact and discrete which means that G must be a finite group. There are 18 abstract point groups. If instead of isomorphism classes, you look at conjugacy classes in the group $GL(3,\mathbb{R})$ of all nonsingular 3×3 real matrices, then you get 32 groups. See Table 1.5 for the abstract point groups. The groups $C_n,\ D_n$, for $n = 1,2,4$ and T, O leave invariant the lattice $\mathbb{Z}^3$. The groups C_n, D_n, for $n = 3,6$ leave invariant a lattice with basis $(1,0,0)$, $(1/2, \sqrt{3}/2, 0)$, $(0,0,1)$. The matrix $-I$ leaves all lattices invariant. More information on these things can be found in Birman [40], Hilbert and Cohn-Vossen [297], Janssen [326], Loeb [424], Lomont [427], Nussbaum [497], and Schwarzenberger [568].

The **crystallographic space groups** are discrete subgroups of the **Euclidean group** $M(3,\mathbb{R})$ of motions of $\mathbb{R}^3$; i.e., the group generated by rotations and translations. The subgroup U of translations in the space group G is isomorphic to $\mathbb{Z}^3$ and G is an extension of U by the finite point group K. There are 219 nonisomorphic space groups. And there are 230 nonconjugate space groups. Schoenfliess, Federov, and Barlow worked this out around 1890, after Sohnke had listed the orientation-preserving space groups. See Schwarzenberger [568, pp. 132–133], for the fascinating history of this episode. It follows that there are only 230 ways to form crystals (although recently quasicrystals have been discovered as we explain in Sect. 1.5.3). Crystallography is based on the study of such groups and

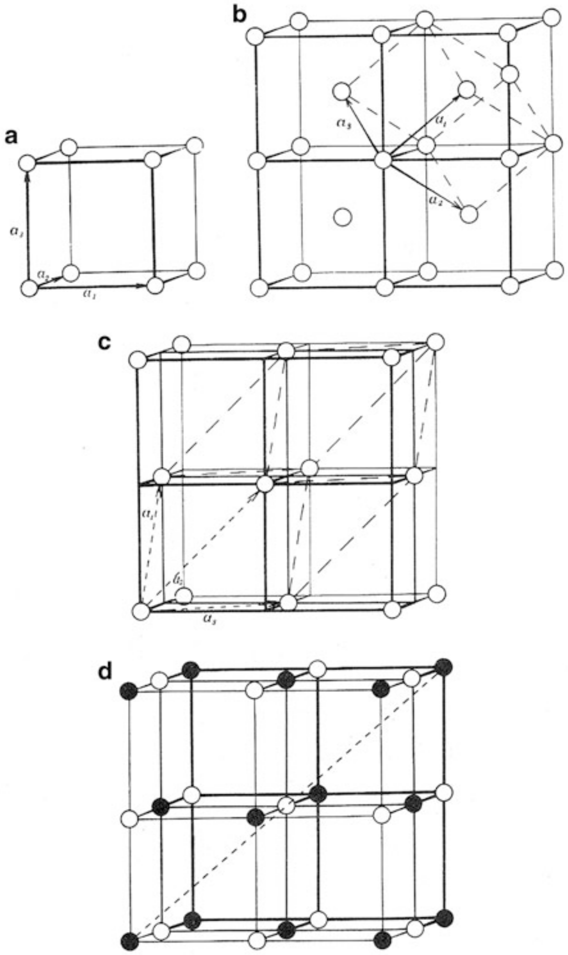


Fig. 1.14 Various crystal lattices: (**a**) simple cubic; (**b**) body-centered cubic; (**c**) face-centered cubic; (**d**) NaCl (From Born and Huang [57, pp. 383–384]. Reprinted by permission of Oxford University Press)

their representations. Chemists know the character tables of these groups quite well. For the quantum mechanics of such a crystal will be ruled by the symmetry group of the crystal, and not the full group of rotations. This will become clearer after we discuss the Schrödinger equation in Sect. 2.1.3.

But we do not want to discuss the representations of finite groups here. Nor do we want to discuss the quantum mechanics that can be derived from knowledge of the character tables of these finite groups leaving crystals invariant. The interested reader could consult some of the references above for these things.

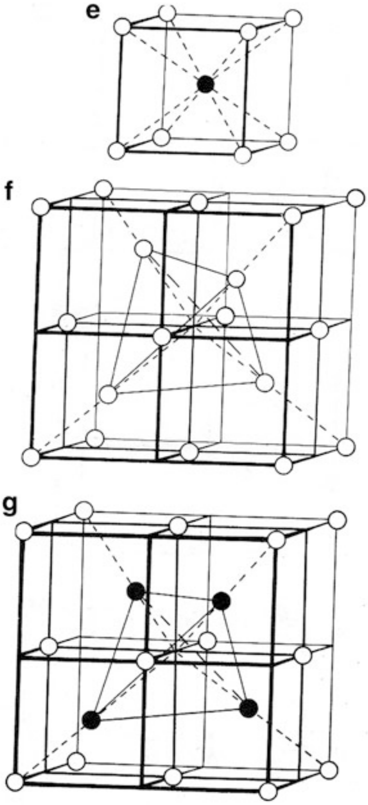


Fig. 1.15 Various crystal lattices: (**e**) CsCl; (**f**) Diamond; (**g**) ZnS (From Born and Huang [57, pp. 383–384]. Reprinted by permission of Oxford University Press)

Table 1.5 Abstract point groups

Cyclic groups $C_n =$ cyclic group with n elements, $n = 1, 2, 3, 4, 6.$
Dihedral groups $D_n =$ group proper rotations (i.e., $\det = 1$) of a regular $n-$gon, $n = 2, 3, 4, 6.$ Tetrahedral group $T =$ group of proper rotations of the regular tetrahedron
Octahedral group $\mathrm{O} =$ group of proper rotations of the cube
Point groups of the second kind add $-\mathrm{I} = -$identity matrix to the group. Write $\mathrm{C}_2 = \{I, -I\}$ $C_4 \times C_2,\ C_6 \times C_2,\ D_2 \times C_2,\ D_4 \times C_2,\ D_6 \times C_2,\ T \times C_2,\ O \times C_2$

Instead we want to show how to use Theorem 1.4.1 to deal with some lattice sums that arise in physics. Given a crystal lattice $\mathbb{Z}v_1 \oplus \mathbb{Z}v_2 \oplus Zv_3$, we can form a positive definite 3×3 matrix

$$Y = \begin{pmatrix} {}^tv_1v_1 & {}^tv_1v_2 & {}^tv_1v_3 \\ {}^tv_2v_1 & {}^tv_2v_2 & {}^tv_2v_3 \\ {}^tv_3v_1 & {}^tv_3v_2 & {}^tv_3v_3 \end{pmatrix} = I[v_1v_2v_3].$$

For any sum over the crystal involving powers of the distance of a vector from the origin we are really computing Epstein zeta functions, since

$$w = a_1 v_1 + a_2 v_2 + a_3 v_3 = (v_1 v_2 v_3) \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix}$$

and $\|w\|^2 = Y[a]$, if $a = {}^t(a_1, a_2, a_3)$ is in $\mathbb{Z}^3$. Many sums of this sort are considered in Born and Huang [57, pp. 248, 385–390, 413], and Ziman [754, pp. 37–41], for example.

Next let us consider the simplest example—the computation of the electrostatic energy of salt (NaCl). The size of this energy has an effect on the solubility of a salt in various solvents, as well as on the hardness and melting point of the crystal (see Højendal [298]). The potential energy of one ion with respect to all others in the lattice is

$$\phi = \frac{\nu \nu_1 e^2}{L_0} \sum (\pm) \frac{L_0}{L} = \frac{\nu \nu_1 e^2}{L_0} \alpha,$$

where νe, $\nu_1 e$ are the charges of the two kinds of ions, L_0 is the length of an edge of the unit cell, L is the distance between two ions, and α is **Madelung's constant**. The sign ± 1 is determined according to charges of the two ions.

A slight generalization of Theorem 1.4.1 leads to a simple way to compute Madelung's constant (see Exercise 1.4.17). This method and variants, some of which use the Fourier expansion of Epstein's zeta function in Chap. 3, have led to many papers in the crystallography literature (cf. the preceding references to Born and Huang or Ziman, plus Emersleben [161, 162], Ewald [169], Sakamoto [553], and Zucker [756], or the multitudes of papers cited by these authors). The series involved in the computation of Madelung's constant for NaCl (salt) is a difference of two Epstein zeta functions evaluated at $s = \frac{1}{2}$. Since Epstein's zeta function of a 3×3 matrix has a pole at $s = \frac{3}{2}$, the Dirichlet series does not converge at $s = \frac{1}{2}$. Thus the series expressing Madelung's constant does not converge absolutely, and a certain amount of care must be taken.

Højendal [298] invented a method that just sums the terms of the original series in a straightforward way. He says that he finds Ewald's method (see Exercise 1.4.17) "still less understandable to persons not acquainted with the highest of mathematics" than Madelung's which involves the Fourier series for Epstein's zeta function (see Chap. 3).

Exercise 1.4.16 (Højendal's Method for Evaluating Madelung's Constant for Salt (NaCl)). Compute α defined above for the NaCl crystal pictured in Figs. 1.14(d) and 1.16. To do this begin by computing $\sum(\pm)L_0/L$ for the eight ions in the smallest possible crystal pictured in Fig. 1.16. You should obtain

$$\frac{3}{1} - \frac{3}{\sqrt{2}} + \frac{1}{\sqrt{3}}.$$

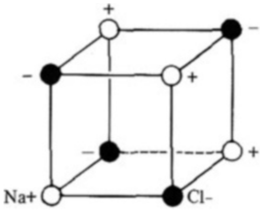


Fig. 1.16 NaCl lattice

Then compute the sum over the next layer of six cubes surrounding this one completely. And continue in this way proceeding by layers similar "to a system of Chinese boxes." We have included Højendal's result so that you can check your program. He had no computer. The nth level has $(2n)^3-1$ ions to deal with. So his ninth division involves summing 8000 terms. Hopefully the reader will use a computer. Our result from the table below is very close to the values found by Madelung and be Emersleben, who obtained 1.747557.

1 division :	$\frac{3}{1}-\frac{3}{\sqrt{2}}+\frac{1}{\sqrt{3}}\cong 1.456030$
2 division :	$\frac{3}{1}-\frac{9}{\sqrt{2}}+\frac{7}{\sqrt{3}}-\frac{3}{\sqrt{4}}+\frac{12}{\sqrt{5}}-$ $-\frac{12}{\sqrt{6}}-\frac{3}{\sqrt{8}}+\frac{6}{\sqrt{9}}-\frac{1}{\sqrt{12}}\cong 0.295739$
3 division :	$\cong -0.004729$
4 division :	$\cong 0.000679$
5 division :	$\cong -0.000221$
	Total 1.747498 ± 0.0005
6 division :	$\cong 0.000100$
7 division :	$\cong -0.000059$
8 division :	$\cong 0.000040$
9 division :	$\cong -0.000028$
estimated for further divisions	$\cong 0.000007$
	Total 0.000060
added to the above	1.747498
we obtain as our most accurate value	1.747558

Exercise 1.4.17 (Ewald's Method or the Method of Theta Functions).

(a) Show that Madelung's constant for NaCl is

$$\sum_{n\in\mathbb{Z}^3-0} \exp\left(\pi i(n_1+n_2+n_3)\right) I\left[n\right]^{-1/2}, \quad I = \text{identity matrix.}$$

This is a special case of Epstein's most general zeta function considered in Epstein [163], for example. The **general Epstein zeta function** is

$$Z(Y,g,h,s) = \sum_{a\in\mathbb{Z}^n-0} Y\,[a+g]^{-s} \exp(2\pi i\,{}^t ha),$$

if $\operatorname{Re} s > n/2,\ Y \in \mathcal{P}_n,\ g,h \in \mathbb{R}^n$.

(b) Define the **general theta function** by

$$\theta(Y,g,h,t) = \sum_{a\in\mathbb{Z}^n-0} \exp\left(-\pi t Y[a+g] + 2\pi i\,{}^t ha\right)$$

for $Y \in \mathcal{P}_n,\ t > 0,\ g,\ h \in \mathbb{R}^n$. Show that theta satisfies the transformation formula

$$\theta(Y,g,h,t) = t^{-n/2}|Y|^{-1/2}\exp(-2\pi i\,{}^t gh)\theta(Y^{-1},h,-g,t^{-1}).$$

(c) Imitate the proof of Theorem 1.4.1 to show that Epstein's general zeta function satisfies the functional equation

$$\begin{aligned}\Lambda(Y,g,h,s) &= \pi^{-s}\Gamma(s)Z(Y,g,h,s)\\ &= |Y|^{-1/2}\ \exp(-2\pi i\,{}^t gh)\Lambda(Y^{-1},h,-g,n/2-s).\end{aligned}$$

Also obtain an expansion of $Z(Y,g,h,s)$ in incomplete gamma functions analogous to the expansion obtained in Theorem 1.4.1.

(d) Use the result of (c) to write a computer program to compute Madelung's constant for NaCl. Compare the results with those of Exercise 1.4.16.

There are other applications of Epstein zeta functions. For example, this function is the simplest case of the zeta functions studied in Minakshisundaram and Pleijel [470] in order to obtain information about the eigenvalues of the Laplacian on compact differentiable manifolds (see also Singer [603]). And minima of Epstein zeta functions for fixed s have been investigated in order to find the best lattice to use in numerical integration among other things (see Delone and Ryskov [128], Fields [179], as well as Sarnak and Strömbergsson [561]). Applications of Epstein zeta functions to quantum-statistical mechanics are considered by Hurt and Hermann [310, Chap. 8]. See also Kirsten and Williams [354].

1.5 Finite Symmetric Spaces, Wavelets, Quasicrystals, Weyl's Criterion

In this section we consider a few topics that were not quite ripe for discussion in 1985—finite analogues of symmetric spaces, wavelets, and quasicrystals, plus a topic of an earlier exercise (Weyl's criterion).

1.5.1 Symmetric Spaces and Their Finite Analogues

Applied mathematicians regularly approximate the continuous with the finite. But it might be argued that one should even start with the finite. D. Greenspan [231] gives many examples and says:

> "It is unfortunate that so many scientists have been conditioned to believe that, say, 10^{30} particles can *always* be approximated well by an infinite number of points. For, indeed, to approximate a physical particle by a mathematical point is to neglect the rich structures of the atom and the molecule, while to approximate 10^{30} objects of any type by an infinite number of such objects is to have enlarged the given set by so great an amount that the given objects are *entirely negligible* in the enlarged set, or, more precisely, in any nondegenerate interval of real numbers, 10^{30} points form a set of measure zero."

The idea of replacing the continuous with the finite occurred to physicists in the 1950s. Quantum mechanics says the state function ϕ for a quantum-mechanical system at energy level E satisfies the Schrödinger equation $L\phi = E\phi$, where L is the Hamiltonian operator. See formulas (1.19) and (2.14), for example. Eugene Wigner had the idea in the 1950s of modelling this with a finite matrix eigenvalue problem. He produced a histogram for the eigenvalues of 197 random symmetric real 20×20 matrices (with normally distributed entries). See Wigner [739]. The result was that the histogram looked like a semi-circle. Thus was random matrix theory born. We say more about this in our second volume on higher rank and general symmetric spaces [667].

Faced with the task of creating finite analogues of symmetric spaces, you may wonder how this can be done. It is possible that you are asking yourself: What is a symmetric space? You have seen one example: Euclidean space $\mathbb{R}^m$. In the next chapter you will investigate another example—the sphere. Chapter 3 concerns a third example the non-Euclidean Poincaré upper half-plane. We have been using the following definition of a **symmetric space** X. The definition says that X is a Riemannian manifold (roughly, a locally Euclidean space with a notion of distance) having a geodesic-reversing isometry at each point; i.e., for each $x \in X$ there is a map $\sigma : X \to X$, preserving the distance between points, fixing x and reversing geodesics through x. Here we are considering noncompact symmetric spaces, except for spheres. We speak of spheres and a little more about symmetric spaces in Chap. 2. We say more about general symmetric spaces in our second volume [667].

To obtain finite analogues of symmetric spaces and to create more examples, it is helpful to define Gelfand pairs. A **Gelfand pair** (G,K) consists of a suitable real Lie group G (which is both a group and a manifold) having a suitable subgroup K such that the algebra $L^1(K\backslash G/K)$ is a commutative algebra under the operation of convolution on G. If (G,K) is a Gelfand pair, then $X = G/K$ gives an example of a symmetric space. We will say more about Lie groups in Chap. 2.

Exercise 1.5.1 (Gelfand's Criterion). Suppose that K is a subgroup of the finite group G and there exists a group isomorphism $\tau : G \to G$ such that $s^{-1} \in K\tau(s)K$ for all $s \in G$. Show that then (G,K) is a Gelfand pair.

For the case of Chap. 1, we were thinking $G=\mathbb{R}^m$, under vector addition, and $K=\{0\}$, but you can also take G to be the Euclidean group and K the subgroup of rotation matrices. You can identify G with the group of matrices of the form $\begin{pmatrix} k & u \\ 0 & 1 \end{pmatrix}$, where $k\in O(m)=O(m,\mathbb{R})$; i.e., ${}^t kk=I$ and the column vector u is in $\mathbb{R}^m$. Elements $g=\begin{pmatrix} k & u \\ 0 & 1 \end{pmatrix}\in G$ act on column vectors $x\in\mathbb{R}^m$ by $gx=kx+u$. Here K is the subgroup fixing the origin; i.e., the subgroup of matrices $\begin{pmatrix} k & 0 \\ 0 & 1 \end{pmatrix}, k\in O(m)$. Thus K can be identified with $O(m)$. This group action leaves invariant the Euclidean distance between points. In Chap. 2, $G=O(3)$ and K is the subgroup fixing the north pole which can be identified with $O(2)$. In Chap. 3, $G=SL(2,\mathbb{R})$, the special linear group of 2×2 real matrices of determinant 1 and K is the subgroup of rotation matrices.

In this computer age it is natural and useful to replace the continuous with some finite approximation. Thus we can replace the real line $\mathbb{R}$ or the circle $\mathbb{R}/\mathbb{Z}$ with the finite circle $\mathbb{Z}/n\mathbb{Z}$ (or with a finite field $\mathbb{F}_q$). This leads to the Fast Fourier Transform or FFT which has revolutionized signal processing. This idea had already occurred to Gauss in 1805 while computing the eccentricity of the orbit of the asteroid Juno. It was not until the paper of Cooley and Tukey in 1965 that the method began to be widely used. Now FFT algorithms are to be found in computer software packages such as Matlab and Mathematica. In this section we shall merely sketch the subject of finite symmetric spaces. References for finite symmetric spaces are Ceccherini-Silberstein et al. [81], Diaconis [132], and my book [668]. The book by Diaconis contains many applications in statistics.

If we carry out this program further, we can replace the orthogonal group of Chap. 2, $G=O(3)=O(3,\mathbb{R})$, with $G=O(3,\mathbb{F}_q)$, where $\mathbb{F}_q$ is a finite field with q elements. Or we can replace $SL(2,\mathbb{R})$ in Chap. 3 with $SL(2,\mathbb{F}_q)$, and then look for subgroups K with the property that (G,K) is a Gelfand pair. We will say more about this later.

Now let's consider a finite analogue of Euclidean space. Suppose that $\mathbb{F}_q$ is a finite field. This implies that $q=p^r$, where p is a prime. Here we assume that $p\neq 2$. If our finite analogue of the real field $\mathbb{R}$ is $\mathbb{F}_q$, then our analogue of finite Euclidean space is $\mathbb{F}_q^m$. Our **finite Euclidean group** $G=M(2,\mathbb{F}_q)$ is the group of matrices of the form $\begin{pmatrix} k & u \\ 0 & 1 \end{pmatrix}$, where the column vector u is in $\mathbb{F}_q^m$ and $k\in O(m,\mathbb{F}_q)$; i.e., k is an $m\times m$ matrix with entries in $\mathbb{F}_q$ such that ${}^t kk=I$. Elements $g=\begin{pmatrix} k & u \\ 0 & 1 \end{pmatrix}\in G$ act on $x\in\mathbb{F}_q^m$ by $g(x)=kx+u$. Here K can be identified with the subgroup of matrices $\begin{pmatrix} k & 0 \\ 0 & 1 \end{pmatrix}$ where $k\in O(m,\mathbb{F}_q)$. Thus we can identify K with $O(m,\mathbb{F}_q)$. We have a finite field valued analogue of the Euclidean **distance** given by

$$d(x,y)=(x_1-y_1)^2+\cdots+(x_m-y_m)^2,\quad \text{for } x,y\in\mathbb{F}_q^m.$$

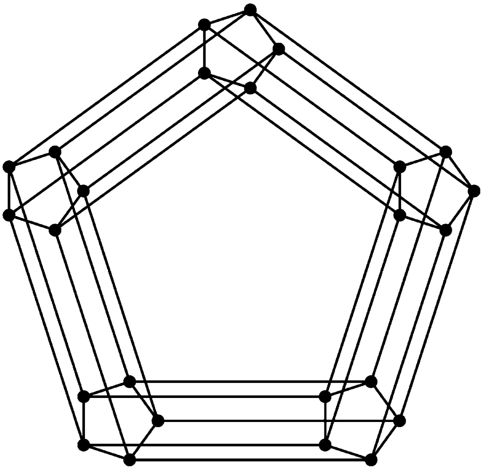


Fig. 1.17 The Euclidean graph $E_5(2,1)$ drawn by Mathematica with the command `ShowGraph [GraphProduct [Cycle [5], Cycle [5]]]` after loading the Combinatorica package

Of course, this "distance" has values in $\mathbb{F}_q$ and thus is not a metric. The group action leaves the Euclidean distance between points invariant.

Exercise 1.5.2. (a) Prove that $d(gx,gy) = d(x,y)$, for all $x,y \in \mathbb{F}_q^m$ and all $g = \begin{pmatrix} k & u \\ 0 & 1 \end{pmatrix}$ acting via $gx = kx+u$, for $k \in O(m,\mathbb{F}_q)$ and $u \in \mathbb{F}_q^m$.

(b) Show that if $G = M(2,\mathbb{F}_q)$ and $K = O(m,\mathbb{F}_q)$, as above, then (G,K) is a Gelfand pair and thus G/K is a finite symmetric space.

Now a new idea occurs to us. One can draw a picture of $\mathbb{F}_q^m$ by attaching a graph to it. Suppose we fix a nonzero element $a \in \mathbb{F}_q$. We define the **Euclidean graph** $E_q(m,a)$ to have vertices the elements of $\mathbb{F}_q^m$ and edges between vertices $x,y \in \mathbb{F}_q^m$ iff $d(x,y) = a$. This is what is called a **regular graph** as it has the same number of edges coming out of each vertex. This number of edges at each vertex in $E_q(m,a)$ is the order of the set

$$S_q(a) = \left\{ u \in \mathbb{F}_q^m \;\middle|\; d(u,0) = a \right\}. \tag{1.28}$$

The number $\left|S_q(a)\right|$ is called the **degree** of the graph $E_q(m,a)$. It can be computed in terms of the quadratic residue symbol. See formula (1.32) for the case that m is even and Terras [668] for the case m is odd. Figure 1.17 shows the graph in the case $E_5(2,1)$. In this special case we get a torus graph—a product of two 5-cycles.

The **adjacency operator** A on a finite graph X acts on functions $f : X \to \mathbb{C}$ by

$$(Af)(x) = \sum_{y \sim x} f(y),$$

where the sum is over all vertices y adjacent to x, meaning that there is an edge between y and x. We use the symbol $y \sim x$ to mean y is adjacent to x. Taking the basis of functions $\chi_a(x) = 1$ for $x = a$ and $\chi_a(x) = 0$ otherwise, the matrix of A is the usual adjacency matrix. For the Euclidean graph $E_q(m,a)$, the **adjacency operator** is

$$(A_a f)(x) = \sum_{\substack{y \in \mathbb{F}_q^m \\ d(y,x)=a}} f(y) = \sum_{u \in S_q(a)} f(x+u), \tag{1.29}$$

where $S_q(a)$ is defined by formula (1.28). The **combinatorial Laplacian** of a finite regular graph X of degree d is defined to be $\Delta = A - dI$, where I is the identity operator; i.e., $If = f$ for all functions on the graph. The definition emphasizes the parallels between graphs and symmetric spaces. See my book [668] for more information.

We can find a complete orthogonal set of eigenfunctions of the adjacency operator of the Euclidean graph $E_q(m,a)$. These are the exponentials appearing in Fourier analysis on $\mathbb{F}_q^m$. To define them, we first need to define the trace of an element of $\mathbb{F}_q$. Suppose $q = p^r$, where p is a prime. Since $x \mapsto x^p$ generates the Galois group of $\mathbb{F}_q$ over $\mathbb{F}_p$, the **trace** of an element $x \in \mathbb{F}_q$ is

$$Tr_{\,\mathbb{F}_q/\mathbb{F}_p}(x) = Tr(x) = x + x^p + x^{p^2} + \cdots + x^{p^{r-1}}.$$

This is the usual trace in field theory, mapping elements of the field $\mathbb{F}_q$ to the subfield $\mathbb{F}_p$. It has the same properties as those of the trace on extensions of number fields. See formula (1.25) in the discussion above of algebraic number theory.

Define the exponential

$$e_b(x) = \exp\left(\frac{2\pi i Tr\left({}^t bx\right)}{p}\right),$$

for $b, x \in \mathbb{F}_q^m$, thinking that b and x are column vectors. Then $\left\{e_b \,\middle|\, b \in \mathbb{F}_q^m\right\}$, is a complete set of characters of $\mathbb{F}_q^m$. This is spelled out in the following exercises. These characters are one-dimensional representations of the group $\mathbb{F}_q^m$ under addition. We will say a little more about group representations in Chap. 2. See also [668].

Exercise 1.5.3. Prove that $e_b(x+y) = e_b(x)e_b(y)$, for all $x, y \in \mathbb{F}_q^m$.

Exercise 1.5.4. (1) Prove that $A_a e_b = \lambda_{a,b} e_b$, where the eigenvalue

$$\lambda_{a,b} = \sum_{u \in S_q(a)} e_b(u). \tag{1.30}$$

(2) Then show that the e_b form a complete orthogonal set of eigenfunctions for the adjacency operator A_a in the inner product space $L^2(\mathbb{F}_q^m)$ consisting of all functions $f : \mathbb{F}_q^m \to \mathbb{C}$.

The eigenvalues $\lambda_{a,b}$ for a and $b \neq 0$ from formula (1.30) satisfy

$$\left|\lambda_{a,b}\right| \leq 2q^{\frac{n-1}{2}}. \tag{1.31}$$

proved using André Weil's estimate for Kloosterman sums. See formula (3.97) for a definition of a Kloosterman sum and [668, Chap. 5], for more details on why the eigenvalues of the Euclidean graphs are essentially Kloosterman sums.

One can show that the degree of these graphs is

$$\left|S_q(a)\right| = q^{m-1} - \varepsilon_q\left((-1)^{\frac{m}{2}}\right) q^{\frac{m-2}{2}}, \text{ when } a \neq 0 \text{ and } m \text{ is even.} \tag{1.32}$$

Here ε_q denotes the **quadratic residue symbol** for the finite field $\mathbb{F}_q$. That means

$$\varepsilon_q(b) = \begin{cases} 1, \text{ if } b \text{ is a square in } \mathbb{F}_q, \\ 0, \text{ if } b = 0, \\ -1, \text{ if } b \text{ is a non-square in } \mathbb{F}_q. \end{cases} \tag{1.33}$$

There is a similar formula for the degree when m is odd. See [668].

A question raised by computer scientists and others about connected finite regular graphs of degree d is: Do the eigenvalues of the adjacency operator satisfy the following bound?

$$\max\{\, |\lambda| \mid \lambda \in spec(A),\ |\lambda| \neq d\} \leq 2\sqrt{d-1}. \tag{1.34}$$

Here $spec(A)$ means the spectrum of A; i.e., the set of all eigenvalues of A. Connected d-regular graphs satisfying this bound are called **Ramanujan graphs**. Infinite sequences of examples of fixed degree with number of vertices approaching ∞ were given by Lubotzky, Phillips, and Sarnak [434], at least for certain degrees. See also the books of Lubotzky [433], and Sarnak [555]. Ramanujan graphs are of interest to computer scientists because they lead to efficient communication networks. The rapid convergence of a random walk on the graph to uniform is behind this fact (at least if the graph is not bipartite). See [668] or Hoory, Linial, and Wigderson [302]. There is an interesting interpretation of the fact that a graph is Ramanujan. It is equivalent to the statement that the zeta function of the graph satisfies the Riemann hypothesis. See [671]. The connection with the conjecture of Ramanujan to be stated in Chap. 3 was the reason for the name of the graphs.

Given the preceding formulas, for an odd prime p, and nonzero $a \in \mathbb{F}_p$, one can see that $E_p(2,a)$ is a Ramanujan graph when $p \equiv 3 (\mathrm{mod} 4)$. Here we use the fact that $\varepsilon_p(-1) = (-1)^{\frac{p-1}{2}}$, as proved in most elementary number theory books. Thus when $p \equiv 3(\mathrm{mod} 4)$, $\left|S_p(2,a)\right| = p+1$, for a not divisible by p. The Ramanujan bound is then $2\sqrt{p}$ and it holds by formula (1.31).

Exercise 1.5.5. What about the Ramanujanicity of $E_p(2,1)$ for $p \equiv 1(\mathrm{mod} 4)$? Check it out using a computer for the first 100 primes.

Next let's say a bit more about the FFT. The beginning of the algorithm is to consider the Fourier transform on $\mathbb{Z}/n\mathbb{Z}$, when $n = rs$. If $f : \mathbb{Z}/n\mathbb{Z} \to \mathbb{C}$, the **Fourier transform** of f is

$$\hat{f}(x) = \sum_{y \in \mathbb{Z}/n\mathbb{Z}} f(y) e^{\frac{-2\pi ixy}{n}}, \ \text{ for } \ x \in \mathbb{Z}/n\mathbb{Z}.$$

It takes n^2 multiplications to compute this transform. But when $n = rs$ we can write the transform as a double sum. For we can write $x = x_2 r + x_1$, with $x_1 \in \mathbb{Z}/r\mathbb{Z}$ and $x_2 \in \mathbb{Z}/s\mathbb{Z}$ and $y = y_1 s + y_2$, with $y_1 \in \mathbb{Z}/r\mathbb{Z}$ and $y_2 \in \mathbb{Z}/s\mathbb{Z}$. This gives setting $w = e^{\frac{-2\pi i}{n}}$:

$$\hat{f}(x) = \sum_{y_2 \in \mathbb{Z}/s\mathbb{Z}} \left(\sum_{y_1 \in \mathbb{Z}/r\mathbb{Z}} f(y_1 s + y_2) w^{x_1 y_1 s} \right) w^{(x_2 r + x_1) y_2}.$$

The inner sum is a Fourier transform over $\mathbb{Z}/r\mathbb{Z}$. The outer one is a Fourier transform over $\mathbb{Z}/s\mathbb{Z}$ of the inner one times $w^{x_1 y_2}$. The computation now takes $sr^2 + rs^2$ operations. This is $n(r+s)$ which is a savings on $n(rs)$. Repeating this idea when n is a power of 2, leads to the original Cooley–Tukey FFT. The FFTs have certainly revolutionized may things from weather prediction to medical imaging. See Terras [668] for more information and references. Other FFTs are discussed by Maslen and Rockmore in [454]. Driscoll and Healy [140] manage to obtain an FFT for the sphere. I am still wondering if there is a way to approximate the Fourier transform on the sphere with that on a quotient of finite orthogonal groups.

1.5.2 Wavelets

Fourier series expand functions as series of complex exponentials or sines and cosines. These functions vanish only at a discrete set of points. Wavelets provide a new approach. The function is expanded in a series of dilations and translations of a "mother wavelet" ψ. The mother wavelet is chosen to have compact support and to be differentiable a finite number of times. In 1988, Ingrid Daubechies showed how to construct such functions (called Daubechies wavelets). See Daubechies [117]. Figure 1.18 shows an example of a Daubechies wavelet as plotted by Mathematica.

There is no simple formula for the Daubechies wavelet function. We will describe how to construct it using the scaling function φ. See Fig. 1.19 for the scaling function corresponding to the wavelet in Fig. 1.18. Now wavelets are used for example in fingerprint analysis, for the JPEG2000 image compression algorithms, in voice recognition software. Here we give only a very brief introduction. More information can be found in Benedetto and Frazier [35], Bratteli and Jorgensen [63], Cipra [90], Ingrid Daubechies [117, 118], Ingrid Daubechies and Lagarias [119], Strang [640], Strang and Nguyen [642], and Strichartz [643].

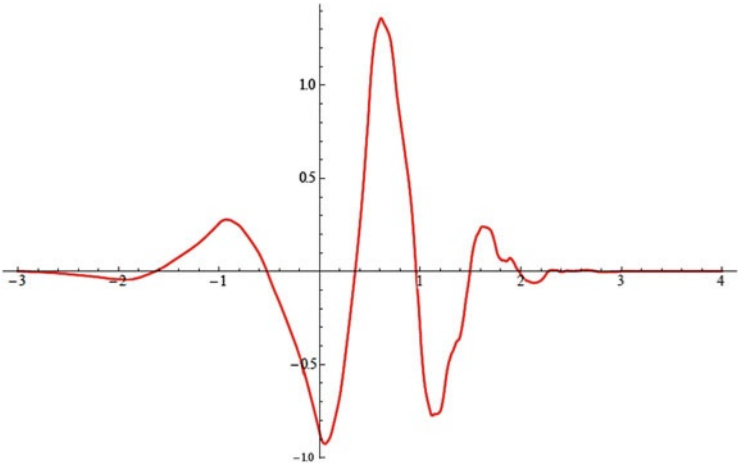


Fig. 1.18 The Daubechies wavelet D4 plotted by Mathematica with the command `Plot[WaveletPsi[DaubechiesWavelet[4], x], {x, -3, 4}, PlotRange -> All, PlotStyle -> {Red, Thick}]`

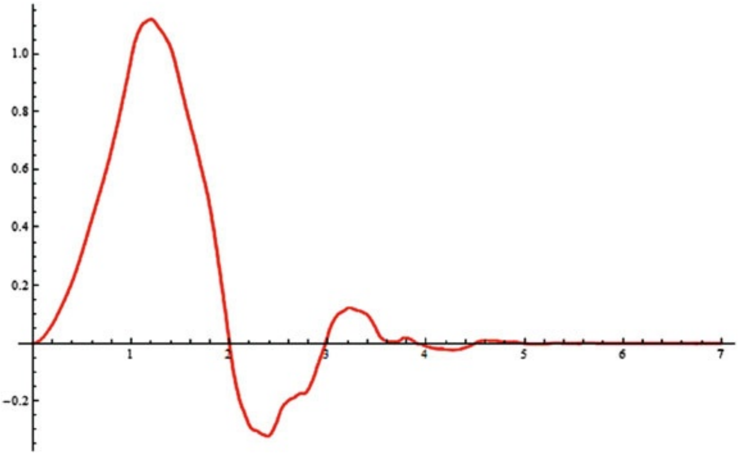


Fig. 1.19 The scaling function for the Daubechies wavelet D4 plotted by Mathematica with the command `Plot[WaveletPhi[DaubechiesWavelet[4], x], {x, 0, 7}, PlotRange -> All, PlotStyle -> {Red, Thick}]`

The scaling function must satisfy a **scaling identity** or dilation equation

$$\varphi(x)=\sum_{k=0}^{N}a_k\varphi(2x-k), \tag{1.35}$$

where the coefficients a_k are chosen very carefully. The number of terms in the sum $(N+1)$ needs to be roughly five times the number of derivatives that you want ψ to have. Restricting φ to the integers turns the scaling identity into an eigenvalue equation for a finite matrix. Once you know φ on $\mathbb{Z}$, the scaling identity determines it on $\frac{1}{2}\mathbb{Z}$. Then use the same method to find φ on $\frac{1}{4}\mathbb{Z}$. Continue inductively to find φ on $2^{-n}\mathbb{Z}$. Then continuity determines φ everywhere. You get the wavelet ψ from the **wavelet formula**

$$\psi(x)=\sum_{k=0}^{N}(-1)^k a_{N-k}\varphi(2x-k). \tag{1.36}$$

The scaling identity implies that we can compute the Fourier transform of the scaling function. We get

$$\hat{\varphi}(y)=p\left(\frac{y}{2}\right)\hat{\varphi}\left(\frac{y}{2}\right), \tag{1.37}$$

where $p(y)$ is the trigonometric polynomial

$$p(y)=\frac{1}{2}\sum_{k=0}^{N}a_k e^{-\pi iky}.$$

Exercise 1.5.6. Prove formula (1.37).

One should choose the coefficients a_k such that $p(0)=1$. Then we find that, upon repeating the process which gave us the formula for $\hat{\varphi}$ and using mathematical induction, we have a formula for the scaling function

$$\hat{\varphi}(y)=\prod_{j=1}^{\infty}p\left(\frac{y}{2^j}\right).$$

The recipe for choosing the coefficients comes from Ingrid Daubechies [117, 118] (see also Benedetto and Frazier [35, p. 38], or Ingrid Daubechies and Lagarias [119]). Once we have the formula for $\hat{\varphi}$, we can use formula (1.36) to obtain the wavelet's Fourier transform:

$$\hat{\psi}(y)=q\left(\frac{y}{2}\right)\hat{\varphi}\left(\frac{y}{2}\right),$$

where q is the trigonometric polynomial

$$q(y)=\frac{1}{2}\sum_{k=0}^{N}(-1)^k a_{N-k}e^{-\pi iky}.$$

Next create the scaled and dilated functions

$$2^{j/2}\psi\left(2^{j}x-k\right)=\psi_{j,k}(x).$$

One then hopes to represent a function f supported in $[0,1]$ by the series $\sum_{j=0}^{\infty}\sum_{k=-\infty}^{\infty}\left\langle f,\psi_{j,k}\right\rangle\psi_{j,k}(x)$. Here the inner product is the usual one for $L^2([0,1])$. There is a method called subband coding that allows the computation of these inner products using just the scaling function coefficients a_k. See J. J. Benedetto and M. W. Frazier [35, p. 46].

I gave an introduction to finite analogs of wavelets and connections with the representations of the finite affine group in [668]. Both Mathematica (Version 8) and Matlab (with the wavelet toolbox) will compute with wavelets.

1.5.3 Quasicrystals

In 2011, Daniel Shechtman won the Nobel prize in chemistry for discovering quasicrystals. He had discovered this structure in 1982 when he chilled a molten mixture of aluminum and manganese. The diffraction pattern of the material was surprising as it showed that the material had a tenfold symmetry. Such a symmetry can be shown to be impossible for classical crystals. See Marjorie Senechal's book [575, p. 6]. Thus in 1992 the International Union of Crystallography changed its definition of crystal. Now it says a crystal is a solid with a "discrete diffraction diagram." The classical definition had required the crystal to have translational symmetry. This will be lacking in quasicrystals. Hundreds of quasicrystals have been created in laboratories. Finally, in 2009, natural quasicrystals, not made in a laboratory, were found.

Even earlier than all this (in 1974) Roger Penrose found nonperiodic tilings of the plane. Ultimately it was shown that these can be made with two tiles. See Gardner [198]. Earlier work had made use of many more tiles. The Penrose tilings were constructed by de Bruijn using a projection from a 5-dimensional lattice. Recently Farris has found a way to visualize such things using sums of projections of eigenspaces of the cyclic permutation matrix in five dimensions. See Farris [175]. I have attempted to imitate this construction in Fig. 1.20. We used the Matlab programs below. The first gives a function associated to the projection of 5-space onto a plane eigenspace of the 5×5 cyclic permutation matrix. The second gives a density plot of the function. Note the appearance of the Golden Ratio $(1+\sqrt{5})/2$.

Matlab Function giving Projection of 5-Spaces onto the Plane

```
function [v] = farris5(x,y)
phi = (1 + sqrt(5))/2;
v = sin(2*pi*x/phi) + sin(2*pi*(x/phi^2 + y/phi))
+sin(2*pi*(−x/phi^2 + y/phi^2))
+sin(2*pi*(−x/phi − y/phi^2)) + sin(2*pi*(−y/phi));
```

Matlab Commands to Plot the Preceding Function

```
% create matrix of farris5 values and density plot of it
clear all
figure
n = input('n =');
for i = 1 : 50*n,
for j = 1 : 50*n,
matfarris5(i,j) = farris5(i/n,j/n);
end
end
pcolor(matfarris5)
```

Exercise 1.5.7. Create a function with a threefold symmetry in a similar way to that used to produce Fig. 1.20. See Farris [175].

So what does harmonic analysis say about all this? We take our discussion from Senechal [575] and Strichartz [643]. See these references for more details. First identify the diffraction diagram as the Fourier transform of the distribution obtained by putting delta functions at the nodes of the quasicrystal. Now we ask: How can we construct a quasicrystal, i.e., a set of nodes such that if we look at the distribution defined by putting delta functions at each node, the result will have a Fourier transform supported on a discrete set?

The idea comes from de Bruijn's method of creating Penrose aperiodic tilings of the plane (see [123]). Let's look at the simplest case. Consider the one-dimensional lattice $\mathbb{Z}^2$ in the plane and rotate it by an angle whose tangent is irrational. Obtain the lattice $\Lambda = A\mathbb{Z}^2$, where $A = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}$, and $\tan\theta$ is irrational. Then take a strip $|y| \leq b$ parallel to the x-axis. Now project the lattice points lying in this strip to the x-axis. That gives an aperiodic tiling of the real line. See Fig. 1.21 for an example known as the Fibonacci tiling of the real line built using the golden ratio $t = \frac{1+\sqrt{5}}{2}$. The red lattice points in Fig. 1.21 have the form $\left((m+tn)/\sqrt{1+t^2}, (n-tm)/\sqrt{1+t^2}\right)$, for $m, n \in \mathbb{Z}$. Those in the horizontal strip of width 4 about the x-axis (marked by green lines in Fig. 1.21) are then projected

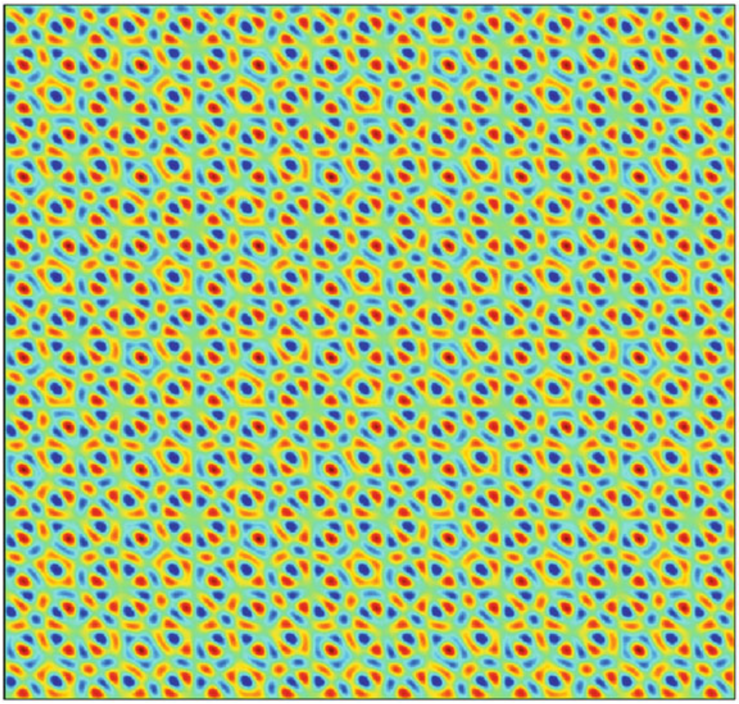


Fig. 1.20 A Matlab plot of the function farris5(x,y) defined above

onto the blue points on the x-axis in Fig. 1.21. We leave it to the reader (and perhaps Google) to uncover the connection between the golden ratio and the Fibonacci numbers.

The distribution corresponding to the set of points

$$n = \left\{(n_1,n_2) \mid n \in \Lambda = A\mathbb{Z}^2, |n_2| \leq b\right\}$$

is

$$L(t) = \sum_{\substack{n=(n_1,n_2)\in\Lambda \\ |n_2|\leq b}} \delta(t-n_1), t \in \mathbb{R}.$$

Why should this have a discrete Fourier transform? If we were to eliminate the restriction on the y-coordinate of the lattice points, in fact, the points n_1 would be uniformly distributed on the x-axis by the Weyl criterion for uniform distribution. It is an exercise using Exercise 1.3.3 of Sect. 1.3 to prove this. Or see the discussion of Kronecker's theorem in the next subsection.

Figure 1.22 shows what Fig. 1.21 would have looked like if we had projected all the lattice points onto the real line rather than restricting ourselves to those with y-coordinates between -2 and $+2$. The blue line looks continuous, though it cannot be. It can only be dense in the real line as it is countable.

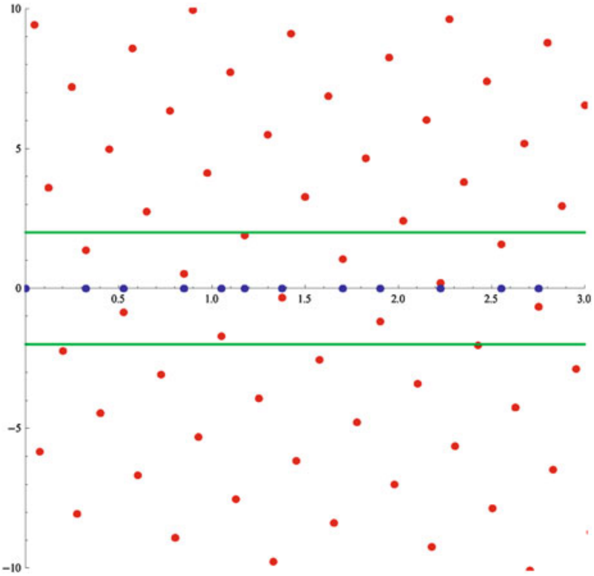


Fig. 1.21 An aperiodic tiling of the real line is given by *blue points*. They are projected from the *red lattice points* that lie between the lines $y = \pm 2$. This is known as the Fibonacci tiling

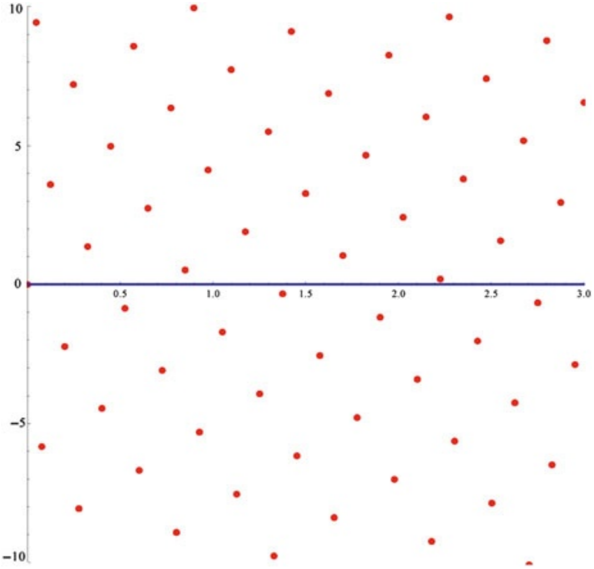


Fig. 1.22 The *blue points* are obtained if you project "all" points in the lattice of Fig. 1.21 onto the real axis, rather than just the points in a strip of width 4 as in Fig. 1.21

The distributional Fourier transform of L is $\sum_{\substack{n\in\Lambda \\ |n_2|\leq b}} e^{-2\pi i n_1 s}$. It is an **exercise** using the Poisson summation formula to show that if $\Lambda = A\mathbb{Z}^2$, where A is a rotation matrix, i.e., ${}^tA = A^{-1}$, we have

$$\sum_{\substack{n=(n_1,n_2)\in\Lambda \\ |n_2|\leq b}} e^{-2\pi i n_1 s} = \sum_{n\in\Lambda} \frac{\sin(2\pi b n_2)}{\pi n_2} \delta(s-n_1).$$

This is a weighted sum of delta functions and thus is supported on a discrete set.

1.5.4 Weyl's Criterion for Uniform Distribution

In Exercise 1.3.3 of Sect. 1.3 we looked at a special case of a criterion due to H. Weyl and this special case really implies that the numbers $\left\{e^{2\pi i a n} \mid n \in \mathbb{Z}\right\}$ or equivalently that the numbers $an(\bmod 1), n \in \mathbb{Z}$, are dense in the circle $\mathbb{R}/\mathbb{Z}$ when a is irrational. This is a result of Kronecker from 1884. Here we wish to discuss uniform distribution and the Weyl criterion which implies it, especially in the two-dimensional case known as the densely wound line in the torus. The result comes from a paper Weyl wrote in 1914. See his collected works [730, Vol. I, pp. 563–599]. See also A. Granville and Z. Rudnick [229], H. Iwaniec and E. Kowalski [320, Chap. 21], T. W. Körner [372, pp. 11–13], or S. J. Miller and R. Takloo-Bighash [467]. In Chap. 3, we will consider non-Euclidean analogues of these results.

Definition 1.5.1. A sequence $\{x_n\}$ in $\mathbb{R}^m/\mathbb{Z}^m$ is **uniformly equidistributed** (with respect to Lebesgue measure on $\mathbb{R}^m/\mathbb{Z}^m$) iff

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} f(x_k) = \int_{\mathbb{R}^m/\mathbb{Z}^m} f(x)dx, \text{ for all continuous } f: \mathbb{R}^m/\mathbb{Z}^m \to \mathbb{C}. \tag{1.38}$$

Theorem 1.5.1 (Weyl's Criterion). *A sequence $\{x_n\}$ in $\mathbb{R}^m/\mathbb{Z}^m$ is **uniformly equidistributed** iff for all $b \in \mathbb{Z}^m - 0$, we have*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} e^{2\pi i\, {}^t b x_k} = 0. \tag{1.39}$$

To sketch the proof in the $\Longleftarrow$ direction (which can be found for $m = 1$ in Körner [372, pp. 11–12]), we just note that formula (1.39) is the case of formula (1.38) with $f(x) = e^{2\pi i\, {}^t bx}$. To use this case to prove formula (1.38) for any continuous function, just uniformly approximate the continuous function by trigonometric polynomials in several variables (as we know we can do this via Fourier series).

The following Exercise is just a rewrite of Exercise 1.3.3 of Sect. 1.3.

Exercise 1.5.8. Suppose that a is irrational and $0 \le c \le d \le 1$. Show that, if the **floor** of $x = \lfloor x \rfloor$ = the greatest integer $<= x$

$$\frac{1}{n}\#\{\ aj - \lfloor aj \rfloor \in [c,d] \mid j = 1,\dots,n\} \to d - c, \ \text{ as } n \to \infty.$$

Then show that this implies Kronecker's theorem which says that the numbers $aj(\operatorname{mod} 1)$, $j \in \mathbb{Z}$, are indeed dense in $\mathbb{R}/\mathbb{Z}$ when a is irrational.

Hint. *Find continuous functions on* $[0,1]$ *uniformly approximating the indicator function* $\chi_{[c,d]}(x) = 1$ *for* $x \in [c,d]$ *and* $\chi_{[c,d]}(x) = 0$ *otherwise. See Körner [372, pp. 12–13].*

The effect of Weyl's criterion is to change an analysis problem to a favorite problem in number theory involving bounds on exponential sums which can often be obtained using results in algebraic geometry such as the Riemann hypothesis for zeta functions of curves over finite fields. It follows from the Weyl criterion that an infinite sequence of points of the form $(j/n, (aj/n) + b)$, $j = -d,\dots,d$, on a line of irrational slope a in the plane induces a sequence of points which become equidistributed on the torus $\mathbb{R}^2/\mathbb{Z}^2$, *as* $d \to \infty$ Thus the line is called "a densely wound line in the torus."

Example 1.5.1 (Densely Wound Line in the Torus). Figures 1.23–1.25 give some idea of what is happening. Of course you cannot really see the difference between a line of rational slope and a line of irrational slope. Later we will look at similar pictures for the hyperbolic plane. In the figures, the torus is represented by a square where you must mentally identify the pair of horizontal sides as well as the pair of vertical sides obtaining first a cylinder and then a doughnut. Thus you really obtain a connected line on the torus or doughnut.

We wrote a Mathematica program to do our torus line pictures. It involves the function:

```
green[x_] := green[x] =
If[(x - Floor[x]) <= .5, x - Floor[x], x - Floor[x] - 1]
```

which finds an integer translate of $x \in \mathbb{R}$ which is in $(-1/2, 1/2]$. We also need the function

```
br[a_, n_, d_] := br[a, n, d] =
ListPlot[Table[{green[j/n], green[a * j/n]}, {j, -d, d}],
AspectRatio-> 1, PlotRange-> {{-1/2, 1/2}, {-1/2, 1/2}},
Background-> LightBlue].
```

Figure 1.23 is a torus line coming from the line with slope 3 produced by the command `br[3,500,500]`. Figure 1.24 is a torus line coming from the line with slope 21/34 produced by the command `br[21/34,1000,80000]`.

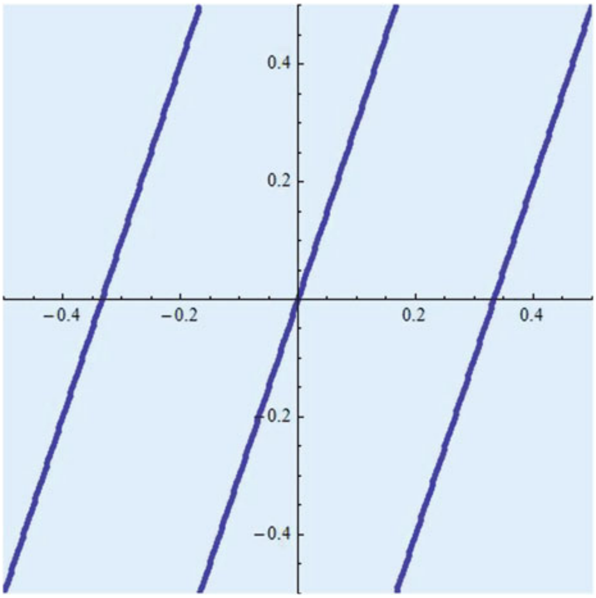


Fig. 1.23 The figure shows points on a line with slope 3

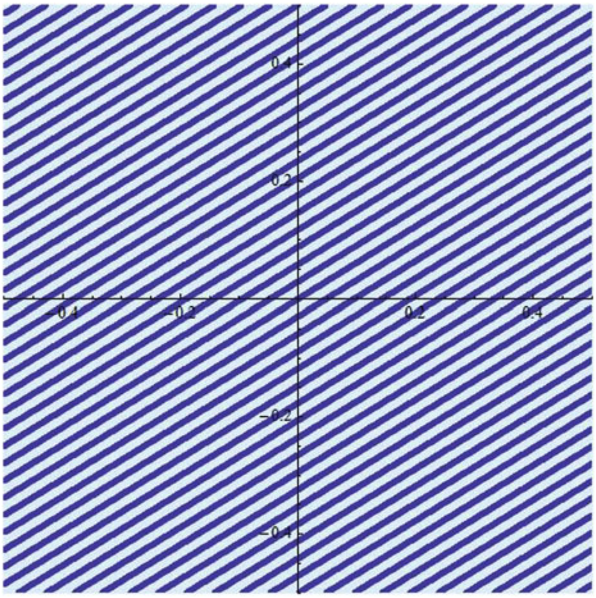


Fig. 1.24 The figure shows points on a line with slope $21/34$

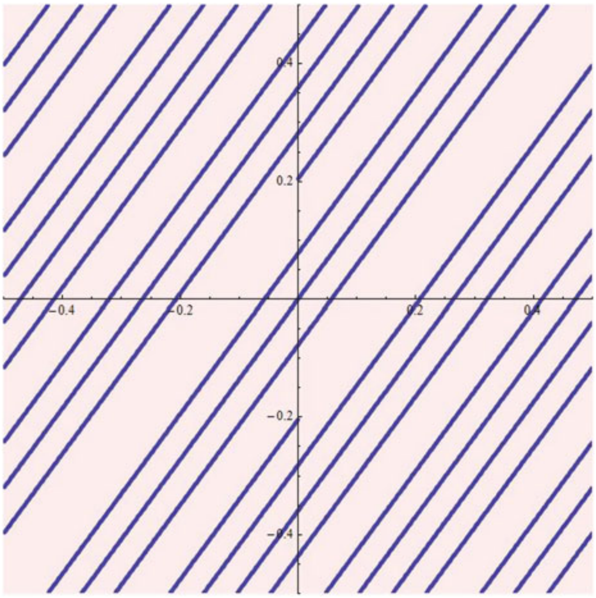


Fig. 1.25 The figure shows some points on a line with slope e/2, where we have changed our command to get a *pink background*

When we used the same command `br[E/2,1000,5000]` (except for a change in background color) for the torus line with slope $E/2$, we got Fig. 1.25.

If we were to increase the last argument of the last variable in `br[E/2,n,d]` (which gives the number of points of spacing $1/n$, on the line through the origin of slope a) to the same value we used for Fig. 1.24, we can make the square totally dark blue in Fig. 1.25. Presumably you do not need to see that picture. However I can certainly take a rational approximation to $e/2$ and produce the same result. It is not really possible to see the difference between $e/2$ and a rational number. In fact `br[2.718/2,1000,80000]` produced a dark blue square as well.

Before leaving this subject, because we are thinking about Kronecker as well as things that are ergodic, fractal, and chaotic (words we have not and may not ever define), we cannot resist looking at an example involving a series of sines that is due to Weierstrass. Kronecker did not approve of this example. In this example, we examine a function that is as non-smooth as one can imagine. In fact, people did not believe that such a function could exist before Weierstrass. Hermite described these functions as a "dreadful plague." Poincaré wrote: "Yesterday, if a new function was invented it was to serve some practical end; today they are specially invented only to show up the arguments of our fathers, and they will never have any other use." Even as late as the 1960s, before "everyone" had a computer fast enough to graph approximations to these things, such examples were viewed as pathological

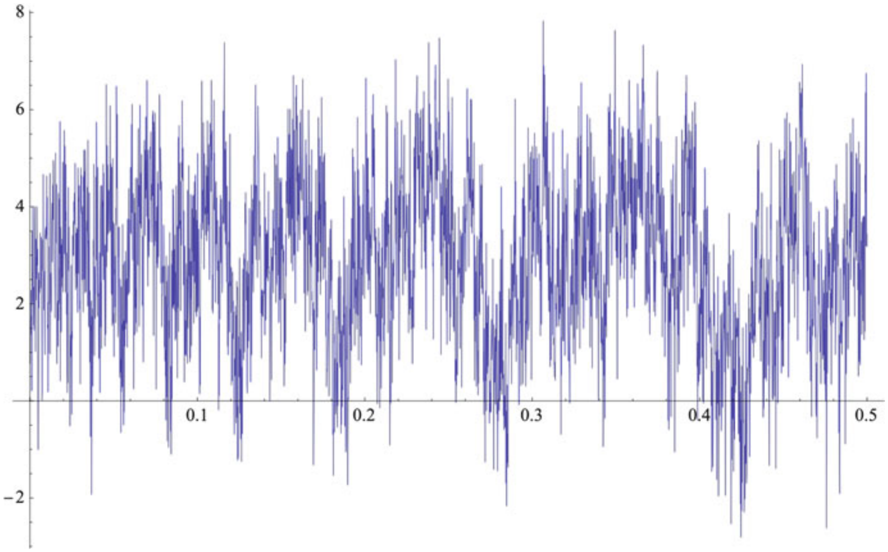


Fig. 1.26 Mathematica plot of the sum $\sum_{k=0}^{3000} 1.5^{(1.9-2)k} \sin\left(1.5^k x\right)$

monsters. Now there are thousands of websites with pictures of approximations of them.

Example 1.5.2. (***Weierstrass Continuous Nowhere Differentiable Function)*** The **Weierstrass function** is defined by an infinite series depending on two parameters λ and s:

$$f(x) = \sum_{k=0}^{\infty} \lambda^{(s-2)k} \sin\left(\lambda^k x\right), \text{ for } \lambda > 1 \text{ and } 1 < s < 2.$$

That $f(x)$ is continuous is an easy consequence of the Weierstrass M-Test for uniform convergence of series of functions. However it can be shown that, if λ is large enough , the derivative $f'(x)$ does not exist at any point x. See Falconer [172, Chap. 11], where it is proved that if λ is large enough the graph of this function is a fractal with box dimension s as defined below.

Definition 1.5.2. Suppose S is a subset of $\mathbb{R}^m$. For any $\varepsilon > 0$, let $N(\varepsilon)$ be the smallest number of m-hypercubes in $\mathbb{R}^m$ needed to cover S. The **box dimension** of S is then defined to be the following limit if it exists:

$$\dim_B S = \lim_{\varepsilon \to 0+} \frac{\log N(\varepsilon)}{\log \frac{1}{\varepsilon}}.$$

The box dimension is often called capacity. It sometimes differs from the Hausdorff dimension which we do not define here. But both definitions of dimension agree on most of the simple examples. One idea of a fractal is that it is a subset of $\mathbb{R}^m$ with non-integer Hausdorff dimension. Mandelbrot coined the term "fractal" in 1977 though many examples were already known at that time. He wanted to obtain a method to study phenomena that were not the usual ones we study in mathematics and said in his 1983 book on the subject "Clouds are not spheres, mountains are not cones, coastlines are not circles, and bark is not smooth, nor does lightning travel in a straight line." Now fractals are of great use in computer graphics. One of the most beautiful fractals is the Mandelbrot set. Sadly the Weierstrass function does not seem so beautiful but it does resemble many phenomena of the modern world such as the rise and fall of the stock market. See Fig. 1.26 for a graph of the Weierstrass function of Example 1.5.2.

Chapter 2
A Compact Symmetric Space: The Sphere

> It is important that you establish a steady, rhythmic tempo while doing the exercises and that you rest as little as possible between them. Also remember not to strain yourself.
>
> —From *Jane Fonda's Year of Fitness and Health/1984, Desk Diary*, Simon and Schuster, NY, 1983. Reprinted by permission.

2.1 Fourier Analysis on the Sphere

2.1.1 The Sphere as a Symmetric Space

> Whenever there is a large earthquake the Earth vibrates for days afterwards. The vibrations consist of the superposition of the elastic–gravitational normal modes of the Earth that are excited by the earthquake.
>
> —From F. Gilbert [212, p. 107].

A (surface or Laplace) **spherical harmonic** is an eigenfunction of the Laplacian on the sphere. These are the analogues of exponentials for Fourier analysis on the sphere. Laplace and Legendre introduced these functions in order to study gravitational theory in the 1780s. Spherical harmonics are necessary for the analysis of any phenomena with spherical symmetry; e.g., earthquakes, the hydrogen atom, and the solar corona. Some of these topics will be discussed later in this section.

Since our treatment of harmonic analysis on the sphere is rather condensed, the reader may want to consult some of the following references for more information: Coifman and Weiss [98], Courant and Hilbert [111], Dym and McKean [147], Erdélyi et al. [164], Lebedev [401], Müller [481], Sugiura [648], Talman [655], Wawrzyńczyk [722], and Vilenkin [704]. For the history of the subject, see Wangerin [716].

Before discussing spherical harmonics, we need to understand something about the geometry of the sphere. This symmetric space is closely related to the **orthogonal group** $O(n)$ of real $n \times n$ matrices U such that ${}^tUU = I$, where ${}^tU$ denotes the transpose of U and I denotes the $n \times n$ identity matrix. The **special orthogonal**

A. Terras, *Harmonic Analysis on Symmetric Spaces—Euclidean Space, the Sphere, and the Poincaré Upper Half-Plane*, DOI 10.1007/978-1-4614-7972-7_2,

© Springer Science+Business Media New York 2013

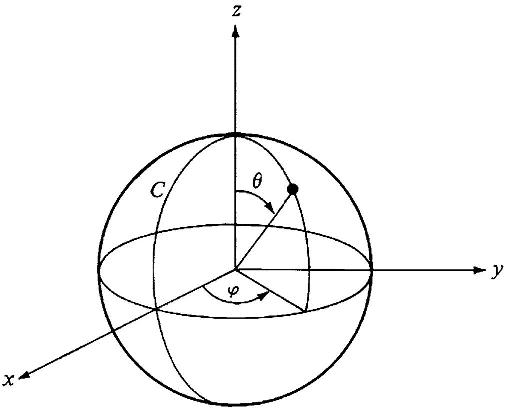


Fig. 2.1 Angular coordinates on the unit sphere

group, $SO(n)$, is the subgroup of matrices U in $O(n)$ such that the determinant of U is one. You can regard U in $SO(n-1)$ as an element of $SO(n)$ by forming

$$\begin{pmatrix} U & 0 \\ 0 & 1 \end{pmatrix} \text{in} SO(n).$$

The group $O(n)$ is a compact group and the sphere is a compact symmetric space, making some of the analysis on it somewhat easier.

Exercise 2.1.1 (The Sphere as a Quotient or Homogeneous Space). Consider the sphere $S^{n-1} = \{x \in \mathbb{R}^n \mid \|x\| = 1\}$. Show that S^{n-1} can be identified with the quotient space $SO(n)/SO(n-1)$, using the preceding identification of $SO(n-1)$ as a subgroup of $SO(n)$.

Hint. *Map the coset $gSO(n-1)$ to the vector ge_n for g in $SO(n-1)$ and $e_n = {}^t(0,\ldots,0,1)$.*

You may also want to read the discussion of the topology of spheres and orthogonal groups in Chevalley [85, pp. 52–67]. In particular, the fundamental (or Poincaré) group of $SO(2)$ is isomorphic to $\mathbb{Z}$, while that of $SO(n)$, $n \geq 3$, has order 2.

From now on we shall consider only the sphere S^2. See the references for the general case. Now the sphere S^2 is a **differentiable manifold.** This means that locally it looks like two-dimensional Euclidean space. To make this precise, we use the usual angular coordinates (θ, φ), $0 < \varphi < 2\pi$, $0 < \theta < \pi$, pictured in Fig. 2.1, to parameterize S^2 except for the semi-circle C through the poles and $(1,0,0)$. A similar coordinate patch can be constructed to cover the rest of the sphere. The equations for the rectangular coordinates (x,y,z) of a point on S^2 in terms of the angular coordinates are

$$\left.\begin{aligned} x &= \sin\theta\cos\varphi \\ y &= \sin\theta\sin\varphi \\ z &= \cos\theta \end{aligned}\right\} \tag{2.1}$$

More information on differentiable manifolds can be found in Yvonne Choquet-Bruhat et al. [87], Helgason [276–278], Loomis and Sternberg [428], Singer and Thorpe [604], or Spivak [614], for example.

Before proceeding further with our discussion of spherical geometry, let us review how to change variables in the Laplacian, using the method of Courant and Hilbert [111, pp. 224–225]. Assume that the substitution mapping (x,y,z) to (u_1,u_2,u_3) is differentiable with differentiable inverse. Let the **Jacobian matrix** of the change of variables be

$$A = \frac{\partial(x,y,z)}{\partial(u_1,u_2,u_3)}. \tag{2.2}$$

Then the **volume element** is

$$dx\,dy\,dz = \|A\|\;du_1\;du_2\;du_3, \quad \|A\| = \text{absolute value of determinant of } A. \tag{2.3}$$

The **Euclidean arc length element** is

$$ds^2 = (dx\;dy\;dz)\begin{pmatrix} dx \\ dy \\ dz \end{pmatrix} = (du_1\;du_2\;du_3)\;{}^tAA\begin{pmatrix} du_1 \\ du_2 \\ du_3 \end{pmatrix}.$$

Thus we obtain

$$ds^2 = \sum_{i,j=1}^{3} g_{ij}du_idu_j \quad \text{where} \quad G = {}^tAA = (g_{ij})_{1\le i,j\le 3}. \tag{2.4}$$

Similarly, one uses the fact that

$$\left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}\right) = \left(\frac{\partial f}{\partial u_1}, \frac{\partial f}{\partial u_2}, \frac{\partial f}{\partial u_3}\right)A^{-1},$$

to see that

$$\left(\frac{\partial f}{\partial x}\right)^2 + \left(\frac{\partial f}{\partial y}\right)^2 + \left(\frac{\partial f}{\partial z}\right)^2 = \sum_{i,j=1}^{3} g^{ij}\frac{\partial f}{\partial u_i}\frac{\partial f}{\partial u_j} \text{ where } G^{-1} = (g^{ij})_{1\le i,j\le 3}. \tag{2.5}$$

To see what happens to the Laplacian in the new coordinate system, one can use the calculus of variations (see Courant and Hilbert [111, Chap. 4]). Suppose that f minimizes the integral

$$J(f) = \int (f_x^2 + f_y^2 + f_z^2)\;dx\;dy\;dz$$

subject to the constraint

$$K(f)=\int f^2\,dx\,dy\,dz=\text{constant}.$$

Then f must satisfy the **Euler–Lagrange equation**

$$f_{xx}+f_{yy}+f_{zz}=\Delta f=\lambda f$$

(see Courant and Hilbert [111, p. 192]). That is, f must be an eigenfunction of the Euclidean Laplacian. Suppose now that we change variables in the integrals J and K. The new constrained minimization problem should lead to a differential equation involving the transformed version of the Laplacian. And this is indeed the case. From (2.3) and (2.5), one obtains

$$J(f)=\int\sum_{i,j=1}^{3}g^{i,j}f_{u_iu_j}\sqrt{|G|}\,du_1\,du_2\,du_3,\qquad\text{with}\qquad |G|=\det G,$$

and

$$K(f)=\int f^2\sqrt{|G|}\,du_1\,du_2\,du_3.$$

The Euler–Lagrange equation for the transformed problem is

$$\sum_i\frac{\partial}{\partial u_i}\sqrt{|G|}\sum_k g^{i,k}\frac{\partial f}{\partial u_k}=\lambda f\sqrt{|G|}.$$

Thus the **Laplacian** in the new coordinate system is

$$\Delta f=f_{xx}+f_{yy}+f_{zz}=|G|^{-\frac{1}{2}}\sum_{i=1}^{3}\frac{\partial}{\partial u_i}|G|^{\frac{1}{2}}\sum_{k=1}^{3}g^{i,k}\frac{\partial f}{\partial u_k}.\tag{2.6}$$

There are other ways to obtain these formulas. See, for example, the following references: Arfken [13, Chap. 2], Yvonne Choquet-Bruhat et al. [87, p. 307], Churchill and Brown [89, pp. 16–19], or Helgason [277, pp. 386–387]).

Now we can show that the sphere is a **Riemannian manifold**, meaning that there is a notion of arc length defined from an inner product on the tangent space to the surface at each point. In order to obtain the arc length element on S^2, we shall use the preceding discussion to write the Euclidean arc length element in spherical polar coordinates. This is done in the next exercise.

Exercise 2.1.2 (Spherical Polar Coordinates). Spherical polar coordinates for a point (x,y,z) in $\mathbb{R}^3$ are defined by

$$x=r\sin\theta\cos\varphi,\ \ y=r\sin\theta\sin\varphi,\ \ z=r\cos\theta,$$

where $0 \le \varphi \le 2\pi,\ 0 \le \theta \le \pi,\ 0 \le r$. Use formulas (2.3), (2.4), and (2.6) above to transform the Euclidean volume, arc length, and Laplacian to spherical polar coordinates.

Answer.

$$\begin{aligned}
&d\mu = dx\,dy\,dz = r^2 \sin\theta\ dr\,d\theta\ d\varphi,\\
&ds^2 = dx^2 + dy^2 + dz^2 = dr^2 + r^2\ d\theta^2 + r^2 \sin^2\theta\ d\varphi^2,\\
&\Delta f = \tfrac{1}{r^2 \sin\theta}\left(\tfrac{\partial}{\partial r}(r^2 \sin\theta \tfrac{\partial f}{\partial r}) + \tfrac{\partial}{\partial\theta}\left(\sin\theta \tfrac{\partial f}{\partial\theta}\right) + \tfrac{\partial}{\partial\varphi}\left(\tfrac{1}{\sin\theta}\tfrac{\partial f}{\partial\varphi}\right)\right).
\end{aligned}$$

It follows from Exercise 2.1.2 that the **element of arc length on the unit sphere** S^2 is

$$ds^2 = d\theta^2 + \sin^2\theta\ d\varphi^2. \tag{2.7}$$

This element of arc length is invariant under $O(3)$ as is the corresponding **volume or area element**:

$$d\mu = \sin\theta\ d\theta\ d\varphi. \tag{2.8}$$

Finally the **Laplacian** on S^2 is

$$\Delta^* = \frac{1}{\sin\theta}\left(\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial}{\partial\theta}\right) + \frac{1}{\sin\theta}\frac{\partial^2}{\partial\varphi^2}\right). \tag{2.9}$$

Exercise 2.1.3. Show that the Euclidean Laplacian $\Delta f = f_{xx} + f_{yy} + f_{zz}$ is invariant under $O(3)$ (i.e., that Δ commutes with rotation). Deduce that the spherical Laplacian (2.9) is also invariant under $O(3)$.

Hint. *Dym and McKean [147, p. 243], have a proof using the Fourier transform.*

Because the sphere is a Riemannian manifold, we can consider geodesics on the sphere. The **geodesic** through two points on the sphere is the curve through these two points that minimizes distance. Airplane pilots know that geodesics in S^2 are great circles; i.e., the intersection of S^2 with a plane through the origin and the two given points on S^2. To see this, suppose that you are given points p and q on S^2. You can rotate S^2 so that both points p and q have φ-coordinate in (2.1) equal to zero. See Exercise 2.1.11 if you do not believe this. Then the distance between p and q on S^2 is

$$\int_p^q \sqrt{d\theta^2 + \sin^2\theta\ d\varphi^2} \ge \int_p^q d\theta = \theta(q) - \theta(p).$$

Thus the great circle route minimizes distance on the sphere (assuming that you go in the proper direction around the circle).

Exercise 2.1.4 (The Sphere Is a Symmetric Space). Show that S^2 is a symmetric space. To do this, it only remains to show that for each point p in S^2 there is a diffeomorphism $f_p : S^2 \to S^2$ that preserves arc length (i.e., f_p is an **isometry** of the Riemannian manifold) and reverses geodesics.

Hint. *At* $p = (0,0,1)$, *use* $(\theta, \varphi) \mapsto (-\theta, -\varphi)$.

Spherical geometry is non-Euclidean because the geodesics cannot be extended indefinitely, which violates Euclid's second postulate. In fact, Euclid's fifth postulate fails as well. And Klein noticed that if you identify antipodal points on the sphere, then any two geodesic lines have a unique point in common. The geometry obtained by identifying antipodal points on the sphere is called **elliptic geometry** and the resulting compact surface is also called the **projective plane**. The term *elliptic* is due to Klein (from a Greek work *elleipein*, to fall short). It is motivated by the following picture. Suppose that two geodesic rays, R_1 and R_2, emanate from the ends of a geodesic segment perpendicular to them both. The distance between the rays R_1 and R_2 will decrease or fall short.

In elliptic or spherical geometry, when A, B, C are the three angles of a geodesic triangle, then the area of the triangle is $A + B + C - \pi$, measuring the angles in radians, of course. This result goes back to Albert Girard (1595–1632) for the sphere. References for non-Euclidean geometry are Coxeter [113] and Hilbert and Cohn-Vossen [297].

Exercise 2.1.5. Prove Girard's formula for the area of a spherical triangle, which was mentioned above.

Note. This could be proved using the Gauss–Bonnet formula (see Singer and Thorpe [604]) and, in fact, it was the first known case of the Gauss–Bonnet theorem.

2.1.2 Spherical Harmonics

It is now possible to make sense of the definition of a surface spherical harmonic, which was given at the beginning of this section. A **surface spherical harmonic** Y is a function $Y : S^2 \to \mathbb{C}$ such that

$$\Delta^* Y = \frac{1}{\sin\theta}\left((\sin\theta \, Y_\theta)_\theta + \frac{1}{\sin\theta} Y_{\varphi\varphi} \right) = \lambda Y \tag{2.10}$$

for some eigenvalue $\lambda \in \mathbb{C}$. The theorem that follows characterizes these spherical harmonics very explicitly in terms of (associated) Legendre polynomials. Legendre had discussed these polynomials before Laplace's work, but Legendre did not manage to publish his results until after Laplace did (see Legendre [404, 405] and Laplace [396]). Spherical harmonics also played a role in Laplace's long treatise on celestial mechanics, as well as in Gauss's work on terrestrial magnetism.

The following exercises give some of the properties of **Legendre polynomials** $P_n(x)$ defined by

$$2^n n! P_n(x) = \frac{d^n}{dx^n}\left((x^2-1)^n\right), \quad \text{for } n = 0, 1, 2, \ldots, \tag{2.11}$$

and **associated Legendre functions** $P_n^m(x)$ defined by

$$P_n^m(x) = (1-x^2)^{m/2} \frac{d^m P_n(x)}{dx^m}. \tag{2.12}$$

Note that some authors replace $(1-x^2)$ by (x^2-1) in formula (2.12). We follow Courant and Hilbert [111] and Arfken [13] rather than Lebedev [401] in our choice of notation.

Exercise 2.1.6 (Legendre Polynomials and Associated Legendre Functions). (a) Show that the associated Legendre functions are solutions of the Sturm–Liouville equation:

$$(1-x^2)u'' - 2xu' + \left(n(n+1) - \frac{m^2}{1-x^2}\right) u = 0.$$

(b) If $n = 0, 1, 2, \ldots$, show that $P_n^m(x) = 0$ for all x, unless $m = 0, \pm 1, \pm 2, \ldots, \pm n$. Note that we can allow m to be negative by writing $(d/dx)^{-1}(d/dx) = \text{identity}$. Show also that

$$P_n^{-m}(x) = (-1)^m \frac{(n-m)!}{(n+m)!} P_n^m(x).$$

(c) Show that for fixed m in $\{0, \pm 1, \pm 2, \ldots\}$, the set

$$\{P_n^m \mid n = |m|,\ |m|+1,\ |m|+2, \ldots\}$$

is a complete orthonormal set for $L^2[-1,\ +1]$.

Hint. *The orthogonality is easy. For completeness when $m = 0$, see Courant and Hilbert [111, pp. 82–83]. In the general case, write $P_{m+k}^m(x) = (1-x^2)^{m/2} f_k^m(x)$. Then $\{f_k^m,\ k = 0, 1, 2, \ldots\}$ is a complete set of orthogonal polynomials for $L^2([-1,+1],(1-x^2)^m)$. This is the weighted L^2-space with inner product given by*

$$(f,g) = \int_{-1}^{+1} f(x)\overline{g(x)}\left(1-x^2\right)^m dx.$$

This means that you must show that f_k^m is orthogonal to x^r when $r = 0, 1, \ldots, k-1$. Integration by parts, using the case $m = 0$, will do the trick.

Exercise 2.1.7 (Integral Formula for Associated Legendre Functions). Show that the associated Legendre P_n^m is represented by the following integral:

$$P_n^m(x) = i^m \frac{\Gamma(n+m+1)}{\pi\Gamma(n+1)} \int_0^\pi \left(x + \sqrt{x^2-1}\cos\varphi\right)^n \cos(m\varphi)\, d\varphi$$

for Re $x > 0,\ m = 0, 1, 2, \ldots.$

Hint. *See Courant and Hilbert [111, p. 505].*

Exercise 2.1.7 has been generalized to symmetric spaces by Harish-Chandra. This will be discussed in later chapters.

Theorem 2.1.1 (Spherical Harmonics).

(a) *The only eigenvalues* λ *of the spherical Laplacian* Δ^* *defined in (2.10) are* $\lambda = -n(n+1),\ n = 0,1,2,\ldots$. *The vector space of eigenfunctions of* Δ^* *corresponding to the eigenvalue* $\lambda = -n(n+1)$ *has dimension* $2n+1$. *A complete orthogonal set of eigenfunctions of* Δ^* *is*

$$Y(\theta,\varphi) = \exp(im\varphi)P_n^m(\cos\theta) \quad \text{for} \quad n = 0,1,2,\ldots, \quad |m| \le n. \tag{2.13}$$

Here P_n^m *is the associated Legendre function (2.12) and the coordinates* (θ,φ) *are defined by (2.1). We call cY, with Y given by formula (2.13), a (surface)* ***spherical harmonic of degree n and order m****, where c is a normalizing constant.*

(b) *Let* $f(x,y,z) = r^n Y(\theta,\varphi)$*, using spherical polar coordinates* (r,θ,φ) *corresponding to the rectangular coordinates* (x,y,z) *in Exercise 2.1.2. Then Y is a surface spherical harmonic satisfying*

$$\Delta^* Y = -n(n+1)Y$$

if and only $f(x,y,z)$ *is a homogeneous harmonic polynomial of degree n. When we say that f is* ***harmonic****, we mean that f satisfies Laplace's equation*

$$\Delta f = f_{xx} + f_{yy} + f_{zz} = 0.$$

Exercise 2.1.8. Prove part (a) of Theorem 2.1.1, using separation of variables on $\Delta^* Y = \lambda Y$ and Exercise 2.1.6.

Note. There are other proofs that $\{\lambda = -n(n+1),\ n = 0,1,2,\ \ldots\}$ are the only possible eigenvalues of Δ^* on S^2. For example, see Dym and McKean [147, pp. 252–253], Kirillov [353, pp. 271–274], or Van der Waerden [691, p. 21].

Exercise 2.1.9. Use separation of variables on $\Delta f = 0$ to prove part (b) of Theorem 2.1.1. Note that Exercise 2.1.2 says that if $f(x,y,z) = R(r)Y(\theta,\varphi)$, then

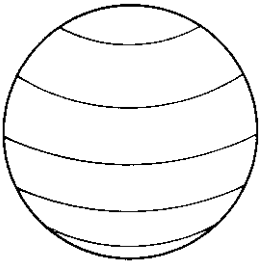


Fig. 2.2 Zero set of $P_n(\cos\theta)$

$$\Delta f = \frac{1}{r^2}(r^2 R'(r))' Y + \frac{R}{r^2}\Delta^* Y \qquad \text{where} \quad ' = \frac{d}{dr}.$$

If $\Delta f = 0$, then $\Delta^* Y = \lambda Y$ and $(r^2 R'(r))' = -\lambda R$, for some constant λ.

You should not be surprised that the eigenvalues of the Laplacian are real since the operator is self-adjoint. The next exercise explains why they are negative.

Exercise 2.1.10 (Why the Eigenvalues of the Laplacian on the Sphere Are Negative). Suppose that f and $\Delta^* f$ are in $L^2(S^2)$. Show that

$$(f, \Delta^* f) \le 0.$$

Use Green's theorem. Then show that the eigenvalues of Δ^* are all negative or zero.

According to part (a) of Theorem 2.1.1, separation of variables in $\Delta^* Y = \lambda Y$ leads from functions $Y(\theta,\varphi)$ of the two angle variables in (2.1) to functions $\exp(im\varphi)P_n^m(\cos\theta)$. If we assume that Y is constant with respect to the angle φ, then $Y = Y(\theta) = P_n(\cos\theta)$. Such a spherical function $Y = Y(\theta)$ is called a **zonal spherical function**. The name results from the fact that the zeros of Y cut the sphere up into zones, as in Fig. 2.2. In general, $P_n(\cos\theta)$ has n distinct zeros in $0 \le \theta \le \pi$ which are positioned symmetrically about $\theta = \pi/2$ (see Fig. 2.2). There is a more group theoretical definition of zonal spherical function, which is developed in the following exercise.

Exercise 2.1.11 (Zonal Spherical Functions on $K\backslash G/K$, where $G = SO(3)$, $K = SO(2)$). Let $G = SO(3)$ and $K = SO(2)$ be considered as a subgroup of G as in Exercise 2.1.1. We know from Exercise 2.1.2 that the sphere can be identified with G/K. Show that the space $K\backslash G/K$ of double cosets KgK, $g \in G$, can be represented by the cosets

$$Kg_\theta K, \quad \text{with} \quad g_\theta = \begin{pmatrix} \cos\theta & 0 & \sin\theta \\ 0 & 1 & 0 \\ -\sin\theta & 0 & \cos\theta \end{pmatrix}, \quad 0 \le \theta < \pi.$$

Table 2.1 Surface spherical harmonics

	$\boldsymbol{Y_n^m(\theta,\varphi)} = (-1)^m\sqrt{\frac{(2n+1)(n-m)!}{4\pi(n+m)!}}P_n^m(\cos\theta)\exp(im\varphi)$
s	$Y_0^0 = \frac{1}{\sqrt{4\pi}}$
p	$\begin{cases} Y_1^{\pm1} = \mp\sqrt{\frac{3}{8\pi}}\sin\theta\ \exp(\pm i\varphi) \\ Y_1^0 = \sqrt{\frac{3}{4\pi}}\cos\theta \end{cases}$
d	$\begin{cases} Y_2^{\pm2} = \sqrt{\frac{15}{32\pi}}\sin^2\theta\ \exp(\pm 2i\varphi) \\ Y_2^{\pm1} = \mp\sqrt{\frac{15}{8\pi}}\sin\theta\ \cos\theta\ \exp(\pm i\varphi) \\ Y_2^0 = \sqrt{\frac{5}{4\pi}}\left(\frac{3}{2}\cos^2\theta - \frac{1}{2}\right) \end{cases}$
f	$\begin{cases} Y_3^{\pm3} = \mp\sqrt{\frac{35}{64\pi}}\sin^3\theta\ \exp(\pm 3i\varphi) \\ Y_3^{\pm2} = \sqrt{\frac{105}{32\pi}}\sin^2\theta\ \cos\theta\ \exp(\pm 2i\varphi) \\ Y_3^{\pm1} = \mp\sqrt{\frac{21}{64\pi}}\sin\theta\ \left(5\cos^2\theta - 1\right)\exp(\pm i\varphi) \\ Y_3^0 = \sqrt{\frac{7}{4\pi}}\left(\frac{5}{2}\cos^3\theta - \frac{3}{2}\cos\theta\right) \end{cases}$

When do two of these cosets coincide? The corresponding points to gK on the sphere are ge_3, where $e_3 = {}^t(0,0,1)$. So if

$$k_\varphi = \begin{pmatrix} \cos\varphi & \sin\varphi & 0 \\ -\sin\varphi & \cos\varphi & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

we find that $k_{-\varphi}g_\theta e_3 = {}^t(\cos\varphi\sin\theta, \sin\varphi\sin\theta, \cos\theta)$. Thus a function on $K\backslash G/K$ is a function only of the angle θ.

Let $A = \{g_\theta, |0 \leq \theta \leq \pi\}$, with g_θ from Exercise 2.1.11. Then the decomposition $SO(3) = KAK$, which follows from the preceding exercise, is the **Euler angle decomposition** of $SO(3)$. See Volume II [667] for a generalization of this decomposition and its application to harmonic analysis on general symmetric spaces.

Exercise 2.1.12. Check Table 2.1 which lists the first few surface spherical harmonics. The table uses the Condon–Shortley [101] convention.

We can use Mathematica to visualize spherical harmonics using density plots on the sphere. See Fig. 2.3 for two such pictures. On the left is that of $\mathrm{Re}\,Y_{14}^7(\theta,\varphi)$ obtained using the Mathematica command:

```
ParametricPlot3D[{Cos[φ]∗Sin[θ], Sin[φ]∗Sin[θ],Cos[θ]},{φ,0,2π},{θ,0,π},
PlotPoints ->300,Mesh->None,ColorFunction->Function[{x,y,z,φ,θ},
Hue[Re[SphericalHarmonicY[14,7,θ,φ]]]],ColorFunctionScaling->
False].
```

On the right is that of $\mathrm{Re}\,Y_{14}^7(\theta,\varphi)$+2$\mathrm{Re}\,Y_7^3(\theta,\varphi)$ obtained using a similar command.

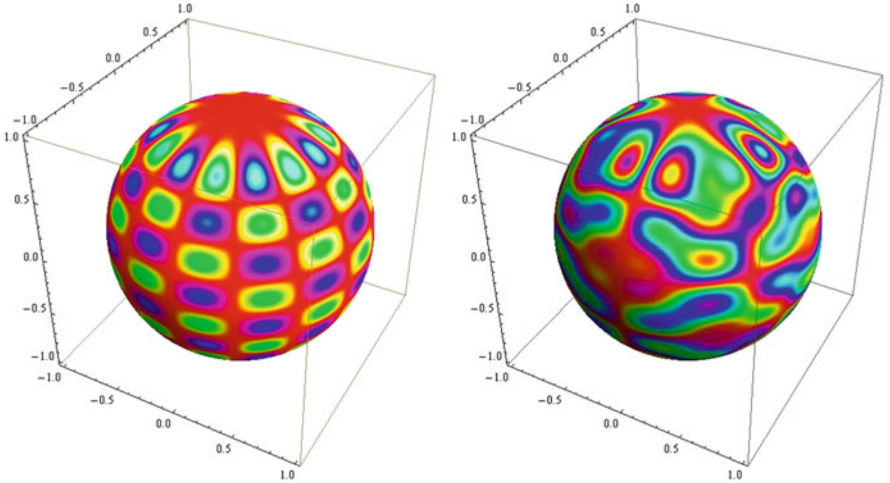


Fig. 2.3 Mathematica density plot for $Y_{14}^{7}(\theta\,,\varphi)$ on the *left* and for $\mathrm{Re}Y_{14}^{7}(\theta\,,\varphi)+2\mathrm{Re}Y_{7}^{3}(\theta\,,\varphi)$ on the *right*

These should be compared with analogous plots for the analogous functions on a unit square in the plane in Fig. 1.4, which we considered in Sect. 1.3.

Exercise 2.1.13. (a) Make some similar plots to those in Fig. 2.3 using the Mathematica command

$$\mathrm{Re}[\mathrm{SphericalHarmonic}Y[m,n,\theta,\varphi]].$$

(b) Solve the wave equation on the sphere with initial condition modelling what would happen if the sphere were struck with a hammer at a point. Then make a Mathematica movie of your solution.

(c) Consider the effect of forcing a sphere to vibrate at some frequency near an eigenfrequency of Δ.

Corollary 2.1.1 (Harmonic Analysis on the Sphere). *Every function* $f: S^2 \to \mathbb{C}$ *with continuous second-order derivatives can be expanded in an absolutely and uniformly convergent series of spherical harmonics. If we take* Y_n^m, $|m| \le n$, *to be an orthonormal basis for* $L^2(S^2, d\mu)$, *with* $d\mu$ *as in Exercise 2.1.2, then*

$$f(\theta,\varphi) = \sum_{n\ge 0}\sum_{|m|\le n} \hat{f}(n,m)Y_n^m(\theta,\varphi),$$

where

$$\hat{f}(n,m) = \int_{S^2} f\,\overline{Y_n^m}\,d\mu.$$

Exercise 2.1.14. Prove the preceding corollary of Theorem 2.1.1.

Hint. *The corollary can be proved from the existence of the Green's function for the problem (see Courant and Hilbert [111, p. 369]) or you might try to relate this problem to Theorem 1.3.1 of Sect. 1.3.*

Exercise 2.1.15 (The Gibbs Phenomenon for the Sphere). Find an analogue of Exercises 1.3.1 and 1.3.2 of Sect. 1.3 for the sphere.

Hint. *There is a discussion of the Gibbs phenomenon as well as summability procedures in Weyl [730, Vol. I, pp. 305–353, 376–389 and Vol. IV, pp. 432–456].*

There are many facts about Fourier series that would be worthwhile to extend to Laplace series of spherical harmonics. For example, it is possible to obtain the analogue of the results of Slepian and Pollak (see Grünbaum et al. [234]). And it is possible to obtain a central limit theorem for the sphere or even $SO(3)$ (see Clerc and Roynette [92]).

Expansions in spherical harmonics are useful in many areas. For example, Backus and Gilbert [19], among others, have used such expansions relative to the vibration of the earth due to large earthquakes to study the structure of the interior of the earth. Some of the flavor of their work can be derived from the following quote.

> The Chilean earthquake of May 22, 1960 excited oscillations of the earth with periods from six to one hundred minutes which were observed continuously on strain gauges for 265 hours after the main shock and at one-minute intervals on a gravimeter for 110 hr. In the Fourier spectra of these records, peaks were observed which corresponded well with the characteristic seismic oscillations computed by Pekeris and Gilbert and MacDonald for certain earth models. Some of the observed peaks, however, differed in period from the theoretical spectral lines, and some of the observed peaks appeared to be double or triple, the components being separated by as much as two minutes and the theoretical line falling approximately at their mean position.
>
> The displacements of observed peaks relative to theoretical frequencies presumably reflect slight errors in the earth model used in the theory and will not be discussed further here. The line splitting cannot be explained thus; we propose to explain it quantitatively as an effect of the diurnal rotation of the earth.

2.1.3 Quantum Mechanics: The Hydrogen Atom

According to the most simplistic version of quantum mechanics (i.e., neglecting spin and relativistic effects), the wave function ψ for the hydrogen atom satisfies the **Schrödinger equation**

$$-\left(\frac{\hbar^2}{2m}\Delta+\frac{e^2}{r}\right)\psi=E\psi, \qquad r=(x^2+y^2+z^2)^{1/2}. \tag{2.14}$$

Here $-e$ is the charge of the electron, $\hbar=$ Planck's constant divided by 2π, Δ is the Laplacian $(\Delta f=f_{xx}+f_{yy}+f_{zz})$, $m=m_e m_p/(m_e+m_p)$ is the reduced mass of the system if m_e is the mass of the electron and m_p the mass of the proton. The eigenvalue E is the **energy level** of the system. Some references for these matters are Biedenharn and Louck [39], Castellan [80], Condon and Odabaşi [100], Courant and Hilbert [111, pp. 341–343], Eyring, Walter, and Kimball [170], Mackey [442, pp. 159–271], Messiah [465, especially pp. 362 and 412], Sommerfeld [610, pp. 200–206], B. Thaller [676], Van der Waerden [691, Chap. 1], Dorothy Wallace and J. BelBruno [713], Weyl [731, pp. 41–70], and Wigner [738].

The eigenvalues E that lie in the discrete spectrum of the Schrödinger operator in equation (2.14) will be the only ones of interest for our discussion. Such eigenvalues are negative (in fact, there is also a continuous spectrum of positive real numbers). Physicists interpret the values of E that lie in the discrete spectrum as **energy levels of bound states of hydrogen**. Spectroscopists going back to Balmer in the 1880s have found various series of lines in the spectrum of hydrogen corresponding to these energy levels. We can obtain some understanding of these spectral lines by finding the discrete spectrum of the Schrödinger operator very explicitly.

We shall use spherical harmonics to separate variables via $\psi(x,y,z) = w(r)Y(\theta,\varphi)$ in equation (2.14) for the purpose of obtaining the discrete spectrum eigenvalues E. Then Exercise 2.1.2 shows that (2.14) becomes

$$\frac{(r^2w')' + (2mr^2/\hbar^2)(E + e^2/r)w}{w} = -\frac{\Delta^* Y}{Y}, \quad \text{where} \quad ' = \frac{d}{dr}.$$

Both sides of this equation must be constant. Thus, by Theorem 2.1.1,

$$\Delta^* Y = -n(n+1)Y, \quad n = 0,1,2,\ldots,$$
$$(r^2w')' + \frac{2mr^2}{\hbar^2}\left(E + \frac{e^2}{r}\right)w = n(n+1)w.$$

Using Exercise 2.1.16 on Laguerre polynomials we find that

$$w(r) = r^n \exp(-cr/(2k)) L_{k+n}^{(2n+1)}(cr/k), \quad k \in \mathbb{Z}, k > n, \ c = 2me^2/(\hbar^2).$$

Furthermore, the corresponding **eigenvalues** are

$$E_k = -\frac{me^4}{2\hbar^2 k^2}. \tag{2.15}$$

Physicists call k the **principal quantum number**, n the **azimuthal quantum number**. If $Y = Y_n^p$, $|p| \le n$, from Theorem 2.1.1, then p is called the **magnetic quantum number**. For each k, there are

$$k^2 = \sum_{n=0}^{k-1}(2n+1) \tag{2.16}$$

linearly independent eigenfunctions of (2.14) with eigenvalue E_k. Physicists call this **k^2-fold degeneracy**. For example, spectroscopists use the letters $(s, p, d, f, q, \ldots)$ to indicate the value of n corresponding to the state of hydrogen. They would say the ground state corresponds to $k = 1$ and is a $1s$ state. The first excited state corresponds to $k = 2$. It is said to be fourfold degenerate because it contains one $2s$ state and three $2p$ states.

Thus for each value k, n of the principal and azimuthal quantum numbers there are $(2n + 1)$ experimentally indistinguishable states of the hydrogen atom. This mathematically explains the **Zeeman effect** in which spectral lines of hydrogen split into an odd number of lines when a magnetic field is switched on. The effect was first observed by Zeeman in 1896.

Note that E_k given by formula (2.15) must equal $h\nu$, where ν is the frequency of a spectral line. And we obtain

$$\nu = Rk^{-2},$$

where R is the Rydberg constant, $R = 2\pi^2 me^4 h^{-3}$. When an atom moves from initial state 1 with principal quantum number k_1 to final state 2 with principal quantum number k_2, the frequency of the associated spectral line is

$$\nu = R(k_2^{-2} - k_1^{-2}).$$

The series of spectral lines of hydrogen observed by Balmer in 1885 has $k_2 = 2$. Thus the Balmer series lines have frequencies

$$\nu = R(2^{-2} - k^{-2}), \quad k = 3, 4, 5, \ldots.$$

These lines are in the visible range. Lyman found a series of spectral lines in the ultraviolet range in 1909 with $k_2 = 1$. The series with $k_2 = 3, 4$, are in the infrared range. The Balmer series is not obtained in the laboratory because it requires the temperature of stellar atmospheres.

The theory just described does not account for the fine structure of the spectrum. Relativistic effects and spin must be considered. Even then, agreement between theory and experiment is not perfect. See Messiah [465, Vol. II, pp. 930–933], for a discussion of the Lamb shift.

There is an additional degeneracy of $(n + 1)^2$ rather than $2n + 1$, which was explained by Fock, who showed that if one formulates the Schrödinger equation for the hydrogen atom as an integral equation, then it is also invariant under $SO(4)$. See Louck and Metropolis [430] for a discussion of a related problem which was motivated by Fock's result.

Spectroscopists daily analyze all sorts of materials—atoms, molecules, crystals, solids, gases. But, of course, the theory becomes much more complicated for the many-body problem. Group theory helps, as many of the references mentioned at the beginning of this discussion show.

Exercise 2.1.16 (Laguerre Polynomials). Set

$$L_p = e^z \frac{d^p}{dz^p}(e^{-z}z^p) = \textit{the pth Laguerre polynomial}.$$

Show that $w(r) = r^n \exp(-cr/(2k)) L_{k+n}^{(2n+1)}(cr/k)$ satisfies

$$(r^2 w')' + \frac{2mr^2}{\hbar^2}\left(E + \frac{e^2}{r}\right) w = n(n+1)w,$$

for E given by equation (2.15), $k \in \mathbb{Z},\ k > n,\ c = 2me^2/(\hbar^2)$. Then show that we have found the only continuous bounded eigenfunctions $w(r)$ for the differential operator involved here.

Hint. *See Arfken [13, Chap. 13], or Courant and Hilbert [111, p. 330].*

We have found a wave function ψ solving equation (2.14) of the form

$$\psi(r,\theta,\varphi) = Y_n^m(\theta,\varphi) r^n \exp(-cr/2k) L_{k+n}^{(2n+1)}(cr/k),$$

where Y_n^m denotes a spherical harmonic as in Theorem 2.1.1 and $L_a^{(b)}$ is the b$^{\text{th}}$ derivative of the Laguerre polynomial L_a. Then $|\psi|^2$ can be interpreted as the probability density of the electron for the various states of the hydrogen atom. Fig. 2.4 from White [733, p. 71], illustrates the first few states.

2.1.4 The Sun's Magnetic Field

Using the measured magnetic field of the sun's surface, the problem is to determine the magnetic field of the corona, assuming that the latter is current-free. The corona itself was not recognized until the last half of the nineteenth century. A new green spectral line was found to appear only in the corona and just above the thin layer next to the solar surface which emits the bright red Balmer line of hydrogen. By 1930, eighteen spectral lines had been discovered in the solar corona. And by the early 1940s it was realized that the green line comes from forbidden transitions in highly ionized metal atoms. The lack of spectral lines of hydrogen and helium in the corona implies that these elements are completely ionized. The corona is mostly hydrogen plasma at one to two million degrees K. There are also cooler regions and warmer regions. Since the corona is too hot to be in thermal equilibrium, there is a solar wind. And observed coronal holes are sources of streams of plasma, which ultimately give rise to terrestrial phenomena such as aurorae and disturbances in radio transmission. A more complete discussion of these basic facts can be found in Altschuler's article in Herman [290, pp. 105–145]. A fascinating account of solar theory can be found in Zirin [755].

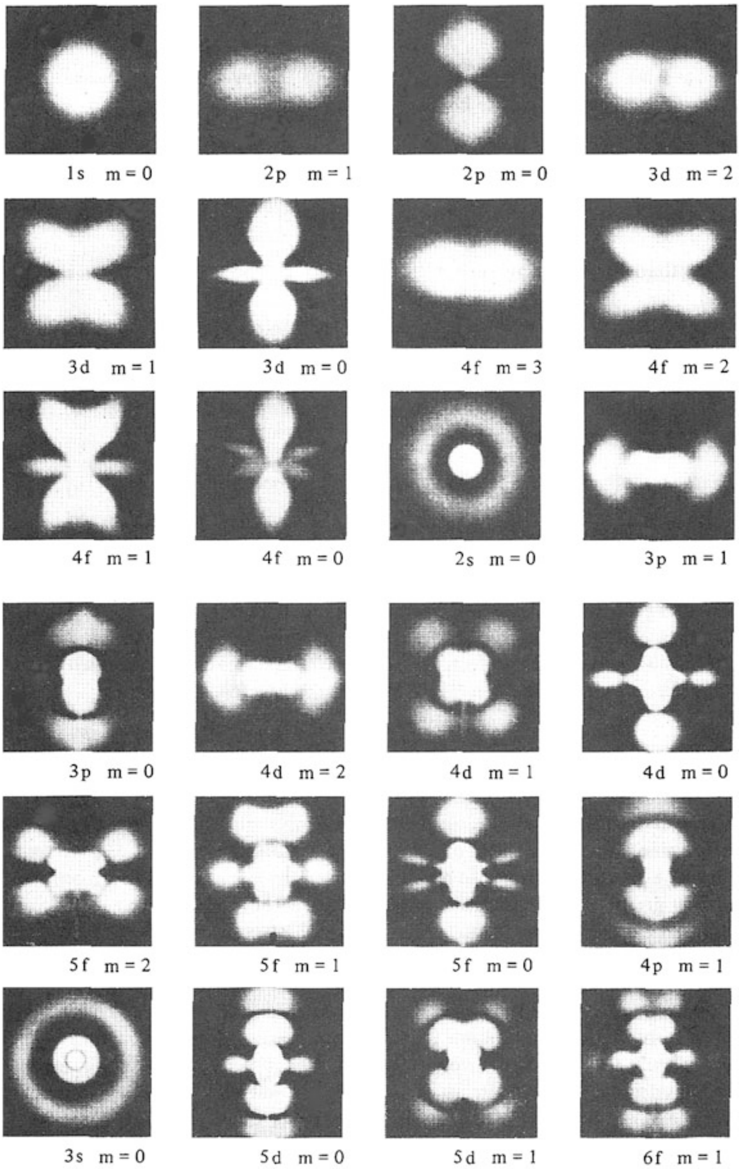


Fig. 2.4 Photographs of the electron cloud for various states of the hydrogen atom as made from a spinning mechanical model. The probability-density distribution $\psi\psi^*$ is symmetrical about the φ-axis, which is vertical and in the plane of paper. (From H.E. White, [733, p. 71]. Reprinted by permission of McGraw-Hill)

The idealized mathematical problem involved in determining the magnetic field of the solar corona using measurements from the solar surface can be posed as follows: find ψ, where

$$\begin{array}{ll} \Delta\psi = 0, & R < \|x\| < R_w, \\ \psi(x) = f(x), & \text{if } \|x\| = R, \text{ the sun's surface,} \\ \psi(x) = 0, & \text{if } \|x\| = R_w, \text{ sufficiently far from the sun.} \end{array}$$

The function $f(x)$ is obtained from the measured magnetic field of the sun's surface. The relation $B = -\text{grad } \psi$ gives the magnetic field itself. The assumption is that there are no magnetic monopoles and thus ψ must be harmonic. So we can use Exercise 2.1.9 and Corollary 2.1.1 to Theorem 2.1.1 to obtain

$$\psi(r,\theta,\varphi) = R \sum_{\substack{n\geq 0 \\ |m|\leq n}} \frac{1}{1-a^{2n+1}} \left(\left(\frac{R}{r}\right)^{n+1} - a^{2n+1} \left(\frac{r}{R}\right)^{n} \right) \hat{f}(n,m) Y_n^m(\theta,\varphi),$$

where $a = R/R_w < 1$ and

$$\hat{f}(n,m) = \int_{S^2} f\, \overline{Y_n^m}\, d\mu.$$

This formula is satisfactory as a solution in a textbook, but the numerical problem remains to be solved. See Altschuler's article mentioned above for a description of the numerical work that led to the illustrations in Fig. 2.5.

Another interesting application of spherical harmonics can be found in Chandrasekhar [83, Chap. VI]. The same sort of picture that gave us an idea of the shape of the probability distribution of the electron in a hydrogen atom should give an idea of the distribution of heat in a uniform sphere. Several applications of spherical harmonics to geophysics can be found in Gilbert [212], Kato [341], Stacey [616], and Trefil [683, pp. 106–111].

It is possible to generalize Theorem 2.1.1 to the unit sphere S^n in $\mathbb{R}^{n+1}$. In fact, spherical harmonics for such higher-dimensional spheres have been of use in quantum mechanics, as have the representations of many higher-dimensional Lie groups. In the Lecture Note-Reprint collection Dyson [148] gives an interesting glimpse into the arguments concerning the choice of symmetry groups for nuclear and particle physics. The compact groups $SU(n)$, $n = 3, 6, 12$, seem to have attracted the most attention. Here $SU(n)$ is the **special unitary group** of $n \times n$ complex matrices U such that ${}^t\overline{U}U = I$ and $|U| = 1$. Some of the papers in the Dyson collection are devoted to showing that the representations of $SU(6)$ are somehow incompatible with those of the Lorentz group $O(3,1)$ which leaves the Maxwell equations invariant. See R. Hermann [291, pp. 144–149] and Talman [655, pp. 186–187] for discussions of Fock's work on applications of four-dimensional spherical harmonics to quantum mechanics. The standard model in particle physics has as its local gauge group $SU(3) \times SU(2) \times SU(1)$. Representations of the group

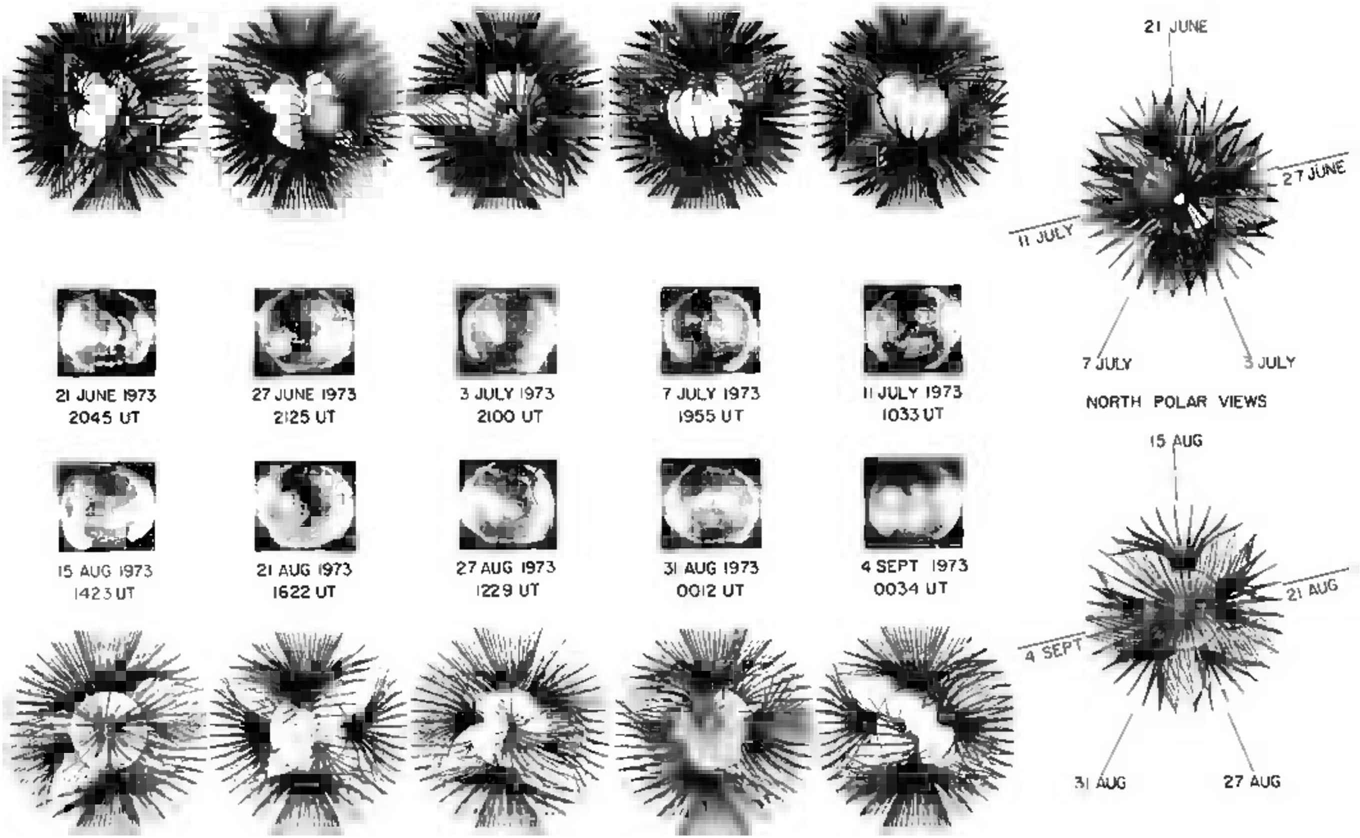


Fig. 2.5 Open magnetic structures on the sun. (From Altschuler's article in Herman [290, p. 125]. Reprinted by permission of Springer-Verlag)

correspond to the fields. There are many books on the subject. We leave it to you to Google them, or see Wikipedia. There is also the Grand Unified Theory, or GUT. This involves $SU(5), SO(10)$ and even the exceptional Lie group E_8. Again there is much on the Internet including some beautiful movies related to E_8.

Hecke and others have made use of these higher-dimensional spherical harmonics in number theory and the kinetic theory of gases (see Hecke [258, pp. 361–373, 849–854]) and Ogg [505, Chap. VI, p. 6]. We shall see in the next section that spherical harmonics also give a new description of harmonic analysis on $\mathbb{R}^n$.

2.1.5 Connections with Group Representations

It is not difficult to connect spherical harmonics with representations of $SO(3)$. Before doing so, let us sketch some of the **basic facts about group representations**. We will be sketchy because we do not intend to emphasize the representation-theoretic point of view. It would be useful for the reader to study these things however, and the following references will provide plenty of food for thought: Barut and Raçzka [28], Boerner [45], Gelfand, Minlos, and Shapiro [205], Gurarie [238], Hamermesh [246], Knapp [356], Mackey [442–444], Maurin [458], Sugiura [648], Talman [655], Van der Waerden [691], Varadarajan [692], Michelle Vergne [697], Vilenkin [704], Wallach [714,715], Warner [717], Wawrzyńczyk [722], Weyl [730–732], Wigner [738], and Williams [741]. Mackey's introduction to Biedenharn and Louck [39] provides a translation of physicists' terminology for mathematicians (or vice versa).

Suppose that H is a **separable Hilbert space** and let $GL(H)$ be the group of invertible continuous linear maps from H to H. We want to consider representations of a topological group G by elements of $GL(H)$. By a **topological group** G we mean a group with a topology such that multiplication $(x,y) \to xy$ and inversion $x \mapsto x^{-1}$ are continuous maps. A topological group is **locally Euclidean** if there is a neighborhood of the identity e in G which is homeomorphic to an open subset of $\mathbb{R}^n$. A **Lie group** is a locally Euclidean topological group with group operations that are infinitely differentiable maps. Most, if not all, of the Lie groups considered here are groups of matrices like $O(n)$.

A **representation** of a topological group G is a pair (T,H) where H is a separable Hilbert space and $T : G \to GL(H)$ is a continuous group homomorphism, using the strong operator topology on $GL(H)$. Actually a stronger definition is made when G is not locally compact (see Kirillov [353, p. 111]). You might wonder why we want to represent groups of matrices by their infinite-dimensional analogues. The answer is that this cannot be avoided if you wish to do Fourier analysis on a noncompact group.

Let (v,w), v, $w \in H$, denote the Hilbert space inner product on H. We say that the representation (T,H) is **unitary** if $(T(g)v, T(g)w) = (v,w)$ for all v,w in H and g in G. The representation (T,H) is said to be **finite-dimensional** if H is finite-dimensional and then dim $H =$ **degree** of the representation.

Suppose that v, w are elements of H. The function $g \mapsto (T(g)v, w) = m_{w,v}(g)$, for $g \in G$, is called a **matrix entry** of the representation (T,H). Many of the special functions which are useful in applied mathematics are either matrix entries of representations of well-known Lie groups or are part of an orthogonal basis for the Hilbert space of such a representation.

The theory of group representations is closely connected with the search for eigenfunctions of differential operators. For if a group G leaves a differential operator L invariant, then G takes eigenfunctions of L into eigenfunctions of L with the same eigenvalue. So G acts on the vector space H of eigenfunctions of L for a fixed eigenvalue λ to give a representation of G.

We say that a representation (T,H) of G is **irreducible** if there is no closed subspace W of H with $\{0\} \subsetneqq W \subsetneqq H$ such that $T(g)W \subset W$ for all $g \in G$.

Two representations (T_1,H_1) and (T_2,H_2) of a Lie group G are said to be **equivalent** if there is a linear bicontinuous bijection $A : H_1 \to H_2$ such that $T_2(g)A = AT_1(g)$ for all g in G. Then A is called an **intertwining operator**. Or even more generally, if A is only continuous linear, it is still called an intertwining operator.

If the group is $\mathbb{R}^m$ under addition, the irreducible unitary representations are one-dimensional and a complete list of inequivalent irreducible unitary representations of $\mathbb{R}^m$ under addition consists of the exponentials

$$\widehat{\mathbb{R}^m} = \left\{ e_a(x) = \exp\left({}^t a x\right), \text{for} \quad x \in \mathbb{R}^m \mid a \in \mathbb{R}^m \right\} \tag{2.17}$$

Exercise 2.1.17 (Schur's Lemma).

(a) Suppose that (T_1,H_1) and (T_2,H_2) are finite-dimensional irreducible representations of G that are inequivalent. If $A : H_1 \to H_2$ is a linear map such that $AT_1(g) = T_2(g)A$, for all g in G, then $A = 0$. Prove this.

(b) Suppose that (T,H) is an irreducible representation of G on a finite-dimensional complex vector space H. If $A : H \to H$ is a linear map such that $AT(g) = T(g)A$, for all g in G, show that there is a complex number c such that $A = cI$, where I is the identity operator; i.e., $Iv = v$, for all v in H.

Note. The spectral theorem for unitary operators allows you to extend the result of the preceding exercise to unitary representations on infinite-dimensional Hilbert spaces.

We say that a finite-dimensional representation (T,H) of G is **completely reducible** if $H = H_1 \oplus \cdots \oplus H_n$, with $T(g)H_i \subset H_i$, for all $g \in G$, provided that the representations (T_i,H_i) of G obtained by restricting $T(g)$ to H_i are irreducible for all $i = 1, \ldots, n$. We say $T = T_1 \oplus \cdots \oplus T_n =$ the **direct sum** of the T_i.

Exercise 2.1.18. Given a basis $e_1, \ldots, e_m$ of the representation space H of a finite-dimensional representation (T,H) of G, you can form the matrix corresponding to $T(g)$. If $e_1, \ldots, e_m$ is an orthonormal basis of H, then this matrix has (i,j) entry $(T(g)e_i, e_j) = m_{j,i}(g)$, for $g \in G$.

(a) Show that two finite-dimensional representations (T_1,H) and (T_2,H) are equivalent if and only if the matrix of T_1 is obtained from the matrix of T_2 by changing the basis of H that is used.

(b) Show that a finite-dimensional representation (T,H) of G is completely reducible if and only if you can find a basis of H which puts the matrix of T in block diagonal form:

$$\begin{pmatrix} T_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & T_n \end{pmatrix}$$

so that the representation determined by the matrix T_i is irreducible for each $i = 1, \ldots, n$.

We will often need to be able to integrate functions on topological groups. The integral involved is an analogue of the Lebesgue integral on $\mathbb{R}^n$ and is supposed to come from a countably additive positive measure on the Borel sets in the group. Such an integral on a topological group G is **right invariant** if

$$\int_G f(x)\,dx = \int_G f(xa)\,dx \quad \text{for all } a \text{ in } G.$$

The invariant integral on a compact Lie group was used by Hurwitz, Schur, and Weyl beginning in the 1890s. Haar proved the existence of a invariant integral on any locally compact topological group in 1933 and thus the integral is called the **Haar integral**. If $G = \mathbb{R}^m$, under addition, then the Haar integral is the standard Lebesgue integral. Proofs of the existence of the Haar integral may be found in Helgason [277, p. 365], Lang [387], Pontryagin [518], or Weil [725], for example.

The invariant integral can be used to show that any irreducible unitary representation of a compact group must be finite-dimensional (see Kirillov [353, p. 135]), for example. And you will need the invariant integral to do Exercise 2.1.20. In fact, integration on topological groups will be a necessary tool for the rest of this book. Of course, our groups will be matrix groups and thus the Haar integral for these groups is not so mysterious.

The right-invariant Haar measure dx is unique up to a positive constant multiple. This means that for each g in G, there is a positive constant $\delta(g)$ (not to be confused with the like-named distribution) defined by

$$\int f(gx)\,dx = \delta(g)\int f(x)\,dx.$$

Then $\delta : G \to \mathbb{R}^+$ is a continuous homomorphism, which is called the **modular function** of G (not to be confused with the modular functions that appear in Chap. 3). Clearly this function relates the right- and left-invariant Haar integrals. In particular, one has

$$\int f(x)\ \delta(x)\ dx = \int f(x^{-1})\ dx.$$

The group G is said to be **unimodular** when $\delta(g) = 1$ for all g in G. This means that right-invariant Haar integrals are also left invariant. Many groups that we shall consider are unimodular. In particular, all compact or abelian groups are unimodular. Most of the Lie groups that we shall consider in this book are what is called "semisimple"; for example, the special linear group of all $n \times n$ real matrices of determinant one. It can be shown that all semisimple Lie groups are unimodular (see Helgason [277, p. 366]). However, not all groups are unimodular. For example, the group of upper triangular matrices with positive diagonal entries is not unimodular if $n \geq 3$ (Exercise).

Exercise 2.1.19. Show that any finite-dimensional unitary representation is completely reducible and that its decomposition into irreducible representations is unique up to equivalence.

Exercise 2.1.20 (The Haar Integral on a Compact Group).

(a) Show that any compact group is unimodular; i.e., right Haar measure = left Haar measure.
(b) Use Haar measure to show that any representation of a compact group is equivalent to a unitary representation.

Suppose that (T_1, H_1) and (T_2, H_2) are representations of G. Define the **tensor product representation** $(T_1 \otimes T_2, H_1 \otimes H_2)$ by

$$(T_1 \otimes T_2)(g)(v_1 \otimes v_2) = (T_1(g)v_1) \otimes (T_2(g)v_2) \quad \text{for} g \in G,\ v_i \in H_i,\ i = 1,2.$$

See the references mentioned above for more information on tensor products. The decomposition of tensor products of representations into direct sums of irreducible representations has much importance for quantum mechanics. For example, if one ignores the interaction between two electrons in the field of a positive nucleus, the Schrödinger operator of the system has eigenfunctions which are products $\psi_1\ \psi_2$ of eigenfunctions ψ_i corresponding to representations T_i of $O(3)$, $i = 1, 2$. And the product $\psi_1 \psi_2$ corresponds to the representation $T_1 \otimes T_2$. The Clebsch–Gordon series breaks $T_1 \otimes T_2$ up into its irreducible components. Thus one can conclude what sort of spectral lines should occur for such a situation. See the references mentioned in Sect. 2.1.3 on quantum mechanics for more details.

There are many other useful topics in representation theory such as induced representations, Frobenius reciprocity, Cartan's theorem on the highest weight, and Weyl's character formula. See the references on group representations for a discussion of these matters.

The systematic study of group representations of finite groups began in the 1890s with work of Frobenius, Schur, and others. In the 1920s, Weyl obtained the irreducible representations of the compact simple Lie groups such as $G = SO(3)$ (see Weyl [730, Vol. II, pp. 543–647]). To do this, Weyl used his formula for the character of a representation T, which is the trace of $T(g), g \in G$. The representations of compact Lie groups G are studied by restricting them to a maximal abelian subgroup A or **torus** in G. Any representation of A decomposes into a direct sum of one-dimensional representations called **weights**. Among the weights there is a highest one and this characterizes the original representation of G, up to equivalence. We have actually seen an example of the theorem on the highest weight in our study of spherical harmonics, but this will not be developed here. Another result called the Borel–Weil–Bott theorem realizes the representations of compact Lie groups on sheaf cohomology groups (see Warner [717]).

The **character** χ_T of a finite-dimensional representation T is defined by $\chi_T(g) = \mathrm{Trace}(T(g))$ for $g \in G$. Define $\hat{G} =$ the **dual** of G to be the set of equivalence classes of irreducible unitary representations of G. For example, since the complete list of irreducible unitary representations is given by formula (2.17), $\widehat{\mathbb{R}^m}$ can be identified with $\mathbb{R}^m$. It is said that $\mathbb{R}^m$ is **self-dual**. If G is compact, describing $\hat{G}$ is equivalent to describing the characters of G.

When infinite-dimensional representations are required, the character is defined as a distribution when possible. If $f \in L^1(G)$, define

$$T(f) = \int_G f(g)T(g)\, dg,$$

which means that for $x, y \in H$, we have

$$(T(f)x, y) = \int_G f(g)(T(g)x, y)\, dg,$$

with $(,) =$ the Hilbert space inner product on H. If H has an orthonormal basis e_i, then we can often define the character of T via

$$\mathrm{Tr}\, T(f) = \sum_i (T(f)e_i, e_i) \quad \text{where } \mathrm{Tr} = \mathrm{Trace},$$

considering the map $f \mapsto \mathrm{Tr}\, T(f)$ as a distribution. When H is finite-dimensional,

$$\mathrm{Tr}\, T(f) = \int_G (\mathrm{Tr}\, T(g)) f(g)\, dg.$$

When G is a "tame" unimodular Lie group the **abstract Plancherel theorem** (or **Fourier inversion theorem**) says there is a measure $d\mu$ on $\hat{G}$ called the **Plancherel measure** such that

$$f(e) = \int_{\hat{G}} \mathrm{Tr}\, T(f)\, d\mu(T),$$

for infinitely differentiable functions f on G with compact support. See Michelle Vergne [697] for a interesting discussion of this subject with many examples. Harish-Chandra obtained the Plancherel measure for real semisimple Lie groups, for which the Plancherel inversion formula does not always involve an integral over all of the dual $\hat{G}$. In Chap. 1 we saw the result for $\mathbb{R}^m$ and $\mathbb{R}^m/\mathbb{Z}^m$. In this chapter we have considered the result for functions on $SO(3)/SO(2)$.

One of the main goals of these volumes is the explicit description of the Plancherel measure—not for functions on the group $GL(n,\mathbb{R})$ but instead for functions on the symmetric space $GL(n,\mathbb{R})/O(n)$. See Volume II (i.e., [667]) for this result and its history.

Formula (2.18) below gives the Plancherel theorem for the group $SO(3)$.

Exercise 2.1.21 (The Irreducible Unitary Representations of $SO(3)$). Let $\{Y_n^m,\ |m| \le n,\ n = 0,1,2,\ \ldots\}$ denote a complete orthonormal set of spherical harmonics as in Theorem 2.1.1. Define a $(2n+1)\times(2n+1)$ matrix $A_n(g)$ for g in $SO(3)$ by

$$Y_n^m(gx) = \sum_{|k|\le n} (A_n)_{m,k} Y_n^k(x).$$

(a) Justify this formula for $A_n(g)$ and then show that $A_n(g)$ defines a representation of $SO(3)$.
(b) Show that $A_n(g)$ is a unitary representation using

$$\int_{S^2} Y_n^m(gx)\overline{Y_n^k(gx)}\,dx = \int_{S^2} Y_n^m(x)\overline{Y_n^k(x)}\,dx.$$

(c) Show that the representation $A_n(g)$ is irreducible.

In fact, the representations $A_n(g)$ form a complete set of inequivalent irreducible unitary representations of $SO(3)$. This means that any function f in $L^2(SO(3))$ with respect to Haar measure has a Fourier series expression, converging in the L^2 norm

$$\left.\begin{aligned} f(g) &= \textstyle\sum_{n=0}^{\infty}(2n+1)\ \mathrm{Trace}\ [\hat{f}(n)A_n(g)], \quad \text{where} \\ \hat{f}(n) &= \textstyle\int_{SO(3)} f(g)\ \overline{{}^tA_n(g)}\ dg, \quad dg = \text{Haar measure on } G. \end{aligned}\right\} \tag{2.18}$$

This is proved, for example, in Dym and McKean [147, pp. 256–261]. Vilenkin [704, pp. 440–457], examines the representations of $SO(n)$ defined in an analogous way to that used for $n = 3$ in Exercise 2.1.21 and shows that if $n > 3$, then these representations do not exhaust all of the irreducible unitary representations of $SO(n)$. Only the class-one representations are obtained in this way. A **class-one representation** (A,H) of G has a vector v in H such that $A(k)v = v$ for all k in K. Here $G = SO(n)$ and $K = SO(n-1)$ (embedded in G as in Exercise 2.1.1).

Formula (2.18) giving Fourier analysis on the group $SO(3)$ is a special case of the **Peter–Weyl theorem** for any compact group (see Weyl [730, Vol. III, pp. 58–75], Pontryagin [518], or Weil [725]). The theorem says that if $\left\{T^\alpha = (m_{ij}^\alpha),\ \alpha \in A\right\}$ is a complete system of irreducible unitary representations of a compact topological group G, then $\sqrt{d_\alpha} m_{ij}^\alpha$ forms a complete orthonormal set in $L^2(G)$, where $d_\alpha =$ degree of T^α. The proof is not hard using the spectral theorem for compact self-adjoint operators (Theorem 1.3.7 of Sect. 1.3). The compact self-adjoint operator used in the proof of the Peter–Weyl theorem is the convolution operator $Tf = g*f$, where g is a fixed nonzero continuous function on G such that $g(x) = g(x^{-1})$ for all x in G. The convolution is taken with respect to Haar measure on G. The eigenfunctions for a fixed eigenvalue of T form a finite-dimensional vector space on which G operates by sending $f(x)$ to $f(xa)$ for $a \in G$. This gives a representation of G. The eigenfunctions for T form a complete orthonormal set in $L^2(G)$ and thus so do the matrix entries of the unitary irreducible representations of G. It follows that the analogue of formula (2.18) replaces A_n by the irreducible unitary representations of G and $(2n+1)$ by the degree of the irreducible representation.

The fact that special functions come from representations often leads to an easier understanding of the myriads of formulas listed in books such as Erdélyi et al. [164, 165].

Exercise 2.1.22 (Addition Formulas for Matrix Entries). Suppose that (T,H) is a finite-dimensional representation of G and that $e_1, \ldots, e_n$ is an orthonormal basis of H. Let the i, j th matrix entry of T be $m_{ij}(g) = (T(g)e_i, e_j)$ for g in G. Here $(,)$ denotes the inner product on H. Show that $T(gh) = T(g)T(h)$ implies the following addition formula for the matrix entries:

$$m_{ij}(gh) = \sum_{k=1}^{n} m_{ik}(h)m_{kj}(g).$$

Exercise 2.1.23 (Addition Formula for Spherical Harmonics). For x, y in S^2, prove that

$$\sum_{|m|\le n} Y_n^m(x)\overline{Y_n^m(y)} = \frac{2n+1}{4\pi} P_n({}^t xy).$$

Note that both sides are unchanged if you replace x by gx and y by gy for g in $SO(3)$. So the right and left sides of the equation can only differ by a constant. Find the constant by setting $x = y$ and integrating over S^2.

2.1.6 Integral Equations for Spherical Harmonics

It is possible to characterize spherical harmonics by integral equations rather than differential equations. This method often simplifies the theory, for the same reason that Green's functions simplify the theory of differential operators (see Sect. 1.3 and the discussion surrounding equation (1.16)). Examples of the method appear in Weyl's work on spherical harmonics (see Weyl [730, Vol. III, pp. 386–399]), Selberg's work on the trace formula (see Selberg [569]), and Harish-Chandra's work on group representations (see Harish-Chandra [253, 254]). The method of integral equations also makes it easier to extend results to finite, p-adic and adelic groups (see Gelbart [200], Godement [213], Macdonald [440], Tamagawa [656]). This should not, however, cause us to forget the differential equations and the contact with applied mathematics.

The following theorem was proved in 1916 by Funk [187] and in 1918 by Hecke [258, pp. 208–214].

Theorem 2.1.2 (The Funk–Hecke Theorem). *Suppose that Y_n is a spherical harmonic of degree n. Let $f : [-1,+1] \to \mathbb{C}$ be continuous. Then*

$$\int_{S^2} f({}^t ux) Y_n(u)\, du = 2\pi Y_n(x) \int_{-1}^{+1} f(t) P_n(t)\, dt.$$

Proof. Since every continuous function on $[-1,+1]$ can be uniformly approximated by zonal spherical functions $P_n(x)$, it suffices to do the following exercises. □

Exercise 2.1.24 (Properties of Spherical Harmonics).

(a) Show that

$$\int_{S^2} Y_m(x) P_n({}^t xy)\, dx = \frac{4\pi}{2n+1} Y_n(y) \delta_{mn}.$$

Here $\delta_{mn} = 0$ if $m \neq n$ and $\delta_{mn} = 1$ if $m = n$.

Hint. *Use the addition formula of Exercise 2.1.23.*

(b) Prove that

$$\int_{-1}^{+1} P_n(x) P_m(x)\, dx = \frac{2\pi}{2n+1} \delta_{mn}.$$

The Funk–Hecke theorem actually characterizes spherical functions. This result is generalized in Helgason [277, p. 439, Cor. 7.4]. And we shall also use this approach in later chapters. In order to generalize the Funk–Hecke theorem, one should view the spherical harmonic Y as a function on $G = SO(3)$ and replace the formula in Theorem 2.1.2 by

$$Y(I) \int_K Y(xky)\, dk = Y(x) \int_K Y(ky)\, dk \quad \text{for all } x, y \text{ in } G, \tag{2.19}$$

where dk is Haar measure on $K = SO(2)$ embedded in G as in Exercise 2.1.1, and $I =$ the identity matrix. Furthermore zonal spherical harmonics P are characterized by the integral equation

$$P(I) \int_K P(xky)\, dk = P(x)P(y) \quad \text{for all } x, y \text{ in } G. \tag{2.20}$$

This characterization of zonal spherical harmonics is due to Gelfand [201] in the general case.

Exercise 2.1.25 (The Group-Theoretical Version of the Funk–Hecke Theorem). Show that formula (2.19) implies the Funk–Hecke theorem by integrating (2.19) times $f(y)$ where $f : K\backslash G/K \to \mathbb{C}$.

Exercise 2.1.26 (The Convolution Theorem for $SO(3)$). Let f,g be integrable functions on $SO(3)$ with respect to Haar measure. Define the **convolution** of f and g by

$$(f*g)(x) = \int_{SO(3)} f(u)g(xu^{-1})\,du \quad \text{for } x \text{ in } SO(3).$$

Define the Fourier transform $\hat{f}(n)$ for $n = 0,1,\ldots$ as in formula (2.18). Show that

$$\widehat{(f*g)}(n) = \hat{f}(n)\hat{g}(n) \quad \text{for all } n = 0,1,2,\ \ldots.$$

Hint. *Use the fact that A_n is a representation.*

Exercise 2.1.27 (The Fundamental Solution to the Heat Equation on the Sphere $S^2 \subset \mathbb{R}^3$). Given the initial heat distribution $f(\theta)$, solve the following initial value problem:

$$\Delta^* u(\theta,\phi,t) = u_t, \quad u(\theta,\phi,0) = f(\theta), \quad t > 0.$$

Here Δ^* is the Laplacian on the sphere given by formula (2.10).

Answer. *$u = G_t * f$, where*

$$G_t(\theta,\phi) = G_t(\theta) = \sum_{n>0} c_n P_n(\cos\theta)\exp[-n(n+1)t].$$

The constants c_n are chosen so that $G_t \to \delta$ as $t \to 0^+$.

Note. G. Watson [720] considers various analogues of the Gaussian distribution for the sphere, in connection with various statistical problems such as that of studying the distribution of the unit normal vectors to the planes of all known comets. Another reference for group theory and statistics is the book by Diaconis [132] which includes an interesting story about testing for uniformity, Fisher, Jeffries, and continental drift on pages 170–171.

2.2 *O(3)* and $\mathbb{R}^3$: The Radon Transform

> It is interesting to consider the method by which algorithms have been "justified" mathematically in this field [computerized tomography]. While this method consists of mathematical

> reasoning in a certain sense, the reasoning is far from rigorous. Approximations are introduced at many steps with only intuition as a guide to the error involved. We do not know of a single instance in which a tomographic algorithm has been justified in a truly rigorous sense. Thus, in contrast to some other workers in this field, we do not feel that one derivation is more rigorous than another, whether it is based on Radon's inversion formula, the Fourier inversion formula or any other foundation.
>
> —From Shepp and Kruskal [588, p. 421].

2.2.1 Harmonic Analysis on $\mathbb{R}^3$ in Spherical Polar Coordinates and Spectral Measures

It is easy to see (cf. Exercise 2.1.1) that the Fourier transform on $\mathbb{R}^n$ commutes with rotation. This leads to a new formulation of harmonic analysis on $\mathbb{R}^n$ in terms of spherical harmonics. For simplicity, we shall consider only the case $n = 3$ here. The generalization to $\mathbb{R}^n$, n arbitrary, can be found in Coifman and Weiss [98] as well as in Stein and Weiss [636]. These results are due to Cauchy and Poisson for radial functions, and Bochner and Hecke in general (see Bochner [43, p. 235], and the article of Stein in J. M. Ash [15, pp. 104–105]).

Exercise 2.2.1 (The Fourier Transform on $\mathbb{R}^n$ Commutes with Rotation). If $g \in O(n)$ and $f : \mathbb{R}^n \to \mathbb{C}$, set $(L(g)f)(x) = f(gx)$ for $x \in \mathbb{R}^n$. Show that $L(g)\hat{f} = \widehat{L(g)f}$, when f is a sufficiently nice function, that the Fourier transform of f exists (as defined in Sect. 1.2 for various classes of functions).

Exercise 2.2.2 (*J*-Bessel Functions). Define the Bessel function of the first kind by the power series

$$J_\nu(z) = \sum_{k=0}^{\infty} \frac{(-1)^k (z/2)^{\nu+2k}}{\Gamma(k+1)\Gamma(k+\nu+1)}, \qquad \text{for } |\arg z| < \pi.$$

(a) Show that $y(z) = J_\nu(z)$ satisfies Bessel's equation:

$$y'' + (1/z)y' + \left(1 - (\nu/z)^2\right) y = 0.$$

(b) Show that $J_\nu(z)$ is represented by the integral formula:

$$J_\nu(z) = \frac{(z/2)^\nu}{\Gamma(1/2)\Gamma(\nu+1/2)} \int_{-1}^{+1} (1-t^2)^{\nu-1/2} \exp(izt)\, dt,$$

when $\operatorname{Re} \nu > -\frac{1}{2}$ and $|\arg z| < \pi$.

(c) Show that $J_{\nu-1}(z) + J_{\nu+1}(z) = \frac{2\nu}{z} J_\nu(z)$.

(d) Prove that

$$J_{n+\frac{1}{2}}(z) = (-1)^n \sqrt{\frac{2}{\pi}} z^{n+\frac{1}{2}} \left(\frac{1}{z}\frac{d}{dz}\right)^n \frac{\sin z}{z}, n = 0,\ 1, 2, \ldots.$$

These functions $J_{n+\frac{1}{2}}$ are often called **spherical Bessel functions**.

Exercise 2.2.3 (More Properties of Bessel Functions).

(a) **Asymptotics.** Show that

$$\begin{aligned} J_\nu(z) &\sim \Gamma(1+\nu)^{-1}(z/2)^\nu, && \text{as } z \to 0; \\ J_\nu(z) &\sim \sqrt{\tfrac{2}{\pi z}} \cos\left(z - \tfrac{\nu\pi}{2} - \tfrac{\pi}{4}\right), && \text{as } z \to \infty. \end{aligned}$$

(b) Show that $J_n(z) = (-1)^n J_{-n}(z)$, when $n = 0, 1, 2, \ldots$. Then show that if ν is not an integer, J_ν and $J_{-\nu}$ are linearly independent.

(c) **Functional Equation.** Prove that $J_\nu(-z) = \exp(\nu\pi i) J_\nu(z)$.

(d) **Addition Formula.** Set $R = (r_1^2 + r_2^2 - 2r_1 r_2\ \cos\theta)^{1/2}$ and prove that

$$J_{\frac{1}{2}}(R) = \sqrt{\frac{2R}{r_1 r_2}} \Gamma\left(\frac{1}{2}\right) \sum_{m=0}^{\infty} \left(m + \frac{1}{2}\right) J_{m+\frac{1}{2}}(r_1) J_{m+\frac{1}{2}}(r_2) P_m(\cos\theta).$$

Theorem 2.2.1 (The Bochner–Hecke Formula). *Suppose that $Y(u)$, $u \in S^2$ is a surface spherical harmonic of degree n (as in Theorem 2.1.1 of Sect. 2.1) and let $f(ru) = g(r)Y(u)$ for $r \in \mathbb{R}^+$, $u \in S^2$, where $g : \mathbb{R}^+ \to \mathbb{C}$, is such that*

$$\int_0^\infty |g(r)| r^2\, dr < \infty.$$

Then the Euclidean Fourier transform as defined in Sect. 1.2 of f (assuming that f is nice enough for the transform to exist), is given by

$$\hat{f}(ru) = \frac{2\pi}{i^n \sqrt{r}} Y(u) \int_{s\in\mathbb{R}^+} g(s) s^{3/2} J_{n+\frac{1}{2}}(2\pi rs)\, ds.$$

Proof. Let dv denote the element of surface area on S^2. Then, using Theorem 2.1.2 of Sect. 2.1, we have

$$\begin{aligned} \hat{f}(ru) &= \int_{s\in\mathbb{R}^+} \int_{v\in S^2} \exp(-2\pi i rs\ {}^t uv) g(s) Y(v) s^2\, ds dv \\ &= 2\pi Y(u) \int_{s\in\mathbb{R}^+} g(s) s^2 \int_{-1}^{+1} \exp(-2\pi rst) P_n(t)\, dt\, ds. \end{aligned}$$

The proof of this theorem is completed in Exercise 2.2.4. □

Exercise 2.2.4. Show that $P_n(x) = \frac{1}{2^n n!} \frac{d^n}{dx^n}(x^2-1)^n$ implies that

$$\int_{-1}^{+1} \exp(-2\pi irst)P_n(t)\,dt = \frac{(-i)^n}{\sqrt{rs}} J_{n+\frac{1}{2}}(2\pi rs).$$

Use integration by parts and Exercise 2.2.2. Then complete the proof of Theorem 2.2.1.

How does the formula of Bochner and Hecke fit into the general scheme of harmonic analysis on symmetric spaces? Why did the Bessel functions suddenly appear out of the blue? To understand this, one must view $\mathbb{R}^3$ as a symmetric space rather than just as an additive group. We explained this at the end of Chap. 1. It leads one to think of the full group of isometries of $\mathbb{R}^3$, namely, the Euclidean group $M(3,\mathbb{R})$ of rigid motions of $\mathbb{R}^3$. If we consider only the orientation-preserving motions G, then G is the semidirect product of the groups T and K, where T consists of translations and is isomorphic to $\mathbb{R}^3$, and $K = SO(3)$. One can view $M(3,\mathbb{R})$ as a group of 4×4 matrices of the form $\begin{pmatrix} A & b \\ 0 & 1 \end{pmatrix}$, $A \in SO(3), b \in \mathbb{R}^3$. The matrix $\begin{pmatrix} A & b \\ 0 & 1 \end{pmatrix}$ sends $x \in \mathbb{R}^3$ to $Ax + b$. Clearly one can identify $\mathbb{R}^3$ with G/K by mapping a coset gK, $g \in G$, to $g0 =$ the result of applying the motion g to the origin 0 in $\mathbb{R}^3$.

Exercise 2.2.5. Do an analogue of the Bochner–Hecke theorem for $\mathbb{R}^2$ for rotation-invariant functions f on $\mathbb{R}^2$.

Answer. *For nice rotation-invariant functions f on $\mathbb{R}^2$, one has*

$$\hat{f}(\rho u) = 2\pi \int_0^\infty f(r)J_0(2\pi\rho r)r\,dr.$$

The connection of Theorem 2.2.2 below with the representations of the Euclidean group (as well as the generalization to $\mathbb{R}^n$) is made in Vilenkin [704, Chap. 11] and Wawrzyńczyk [722]. The case of $\mathbb{R}^2$ is discussed in Dym and McKean [147, pp. 263–273]. See also Helgason [277, pp. 402–403], and Talman [655, Chap. 12].

Exercise 2.2.6 (The Eigenfunctions of the Euclidean Laplacian in Spherical Polar Coordinates). Use Exercise 2.1.2 of Sect. 2.1 to show that if $\Delta f = f_{xx} + f_{yy} + f_{zz} = \lambda f$, and $f(ru) = g(r)Y(u)$ for $r \in \mathbb{R}^+$, $u \in S^2$, and Y is a surface spherical harmonic of degree n, then

$$g(r) = \frac{1}{\sqrt{rt}} J_{n+\frac{1}{2}}(2\pi rt),\ \lambda = (2\pi t)^2.$$

Theorem 2.2.2 (Harmonic Analysis on $\mathbb{R}^3$ in Spherical Polar Coordinates). *Let $e_{n,m,t}(ru) = 2\pi Y_n^m(u)(rt)^{-1/2}J_{n+\frac{1}{2}}(2\pi rt)$ for $r \in \mathbb{R}^+$, $u \in S^2$, where $\{Y_n^m \mid n = 0,\ 1,2,\ldots,|m| \le n\}$ denotes a complete orthonormal set of surface spherical* *harmonics of degree n (as in Theorem 2.1.1 of Sect. 2.1). Then any f in* $L^2(\mathbb{R}^3)$ *has a Fourier expansion (converging in the* L^2*-norm) of the form*

$$f(x) = \sum_{n\geq 0}\sum_{|m|\leq n}\int_{t>0}\hat{f}(n,m,t)e_{n,m,t}(x)t^2\,dt,$$

where

$$\hat{f}(n,m,t) = \int_{\mathbb{R}^3} f(y)\overline{e_{n,m,t}(y)}\,dy.$$

Proof. It suffices to assume that $f(ru) = g(r)Y_n^m(u)$ for $r \in \mathbb{R}^+$, $u \in S^2$. Then we must show that

$$g(r) = (2\pi)^2\int_{s>0}\left(\int_{t>0} g(t)\frac{1}{\sqrt{st}}J_{n+\frac{1}{2}}(2\pi st)t^2\,dt\right)\frac{1}{\sqrt{rs}}J_{n+\frac{1}{2}}(2\pi rs)s^2\,ds. \quad (2.21)$$

This is done in the following exercise. □

Exercise 2.2.7. (a) Prove formula (2.21) by writing down the inversion formula for the ordinary Fourier transform (from Theorem 1.2.1 of Sect. 1.2) in spherical polar coordinates. Then use Theorem 2.2.1 to evaluate the inner Fourier transform. Finally, Theorem 2.1.2 of Sect. 2.1 and Exercise 2.2.4 should be used to complete the proof.

(b) Show that formula (2.21) is equivalent by change of variables to a special case of **Hankel's inversion formula**: (assuming $\nu > -\frac{1}{2}$ and $r > 0$) :

$$f(r) = \int_0^\infty yJ_\nu(yr)\int_0^\infty xJ_\nu(xy)f(x)\,dx\,dy. \quad (2.22)$$

Hankel's inversion formula (2.22) gives the spectral resolution of the singular Sturm–Liouville operator

$$(Lf)(x) = \frac{1}{x}\left(-(xf')' + \frac{\nu^2}{x}f\right) \quad \text{for } x \in (0,\infty) \quad (2.23)$$

This spectral resolution can be derived from a formula of Stieltjes, Stone, Kodaira, and Titchmarsh, which is itself a corollary of the Von Neumann Spectral Theorem for unbounded operators. Let us summarize the theory briefly, following Lang [387], Reed and Simon [540, Chap. 7], Stakgold [618, 619], and some unpublished notes of J. Korevaar. More details can be found in these references as well as Dunford and Schwartz [145], Gelfand and Vilenkin [207], Levitan and Sargsjan [415], Maurin [459], Naimark [490], Titchmarsh [679], Weyl [730, Vol. I, pp. 195–297], and Yosida [748]. In particular, Dunford and Schwartz [145, Vol. II, pp. 1333–1392, 1532–1533], provides an extremely careful discussion of spectral theory for singular differential operators, including the Bessel operator (2.23).

We want to consider linear, possibly unbounded, operators $T : D \to V$, where D is a dense subspace of a Hilbert space V. Define

$$D^* = \{y \in V \,|\, \text{there is a } z \text{ in } V \text{ with } (Tx,y) = (x,z), \text{ for all } x \in D\}.$$

Define T^* = the **adjoint** of T by $(Tx,y) = (x,T^*y)$. We say that T is **self-adjoint** if $D = D^*$ and $T = T^*$. The **spectrum** $\sigma(T)$ of the operator T is defined to be

$$\sigma(T) = \left\{\lambda \in \mathbb{C} \;\middle|\; (T-\lambda I)^{-1} \text{ does not exist as a bounded linear operator on } V\right\}.$$

The **point or discrete spectrum** of T (or the set of **eigenvalues** of T) consists of $\lambda \in \mathbb{C}$ such that $(T - \lambda I)$ is not one-to-one. The **continuous spectrum** of T is the set of $\lambda \in \mathbb{C}$ such that $(T - \lambda I)$ is 1 - 1 and D = range $(T - \lambda I)$ is dense in V but $(T - \lambda I)^{-1}$ is an unbounded linear operator with domain D.

The **von Neumann Spectral Theorem** says that if T is any self-adjoint operator (bounded or not), there is a Stieltjes integral representation

$$I = \int dE_\lambda, \quad T = \int \lambda dE_\lambda, \tag{2.24}$$

for some family of projection-valued measures dE_λ (cf. Reed and Simon [540, Chap. 7]). The integrals are over the spectrum of T, which is real, because T is self-adjoint. The spectral theorem implies that for polynomials $p(\lambda)$, $p(T) = \int p(\lambda) dE_\lambda$. It is possible to define $f(T)$ for continuous functions f on the spectrum of T and then this same formula holds upon replacing p by f.

In order to obtain the formula of Stieltjes, Stone, Kodaira, and Titchmarsh one must be able to compute the **Green's function** or **resolvent kernel** $G(\lambda;x,y)$ defined by

$$(T-\lambda I)^{-1} v(x) = \int G(\lambda;x,y) v(y)\, dy. \tag{2.25}$$

For if we have a convergent integral of the form

$$f(\mu) = \int_{-\infty}^{+\infty} (\lambda - \mu)^{-1} g(\lambda) d\lambda,$$

then at any point c of continuity of $g(\lambda)$, we have

$$\begin{aligned} &\tfrac{1}{2\pi i} \lim_{\varepsilon \to 0+} \left(f(c+i\varepsilon) - f(c-i\varepsilon)\right) \\ &\quad = \lim_{\varepsilon \to 0+} \left\{ \tfrac{1}{\pi} \int_{-\infty}^{+\infty} \tfrac{\varepsilon}{(\lambda-c)^2+\varepsilon^2}\, g(\lambda)\, d\lambda \right\} = g(c). \end{aligned}$$

This is Poisson's integral formula for a half-plane (see Ahlfors [3, p. 171]). It follows that the spectral measure for the operator T is given by the **formula of Stieltjes, Stone, Kodaira, and Titchmarsh:**

$$2\pi i \frac{dE_\lambda}{d\lambda} = (T-(\lambda+i0)I)^{-1} - (T-(\lambda-i0)I)^{-1}, \tag{2.26}$$

assuming that E_λ is well-behaved. This formula is proved in Dunford and Schwartz [145, Vol. II], Lang [387, pp. 412–413], and Reed and Simon [540, p. 237]. Stieltjes [639] found the result in 1894. We will usually refer to formula (2.26) as the **Kodaira–Titchmarsh formula** for brevity.

In order to apply formula (2.26), we need a way to compute the Green's function $G(\lambda;x,y)$. We use the method given in Stakgold [618, 619]. Consider a **Sturm–Liouville operator which is singular** at a and b:

$$Lf = \frac{1}{w}\left(-(pf')' + qf\right), \quad a<x<b. \tag{2.27}$$

Here "singular" means that either the interval is infinite or the functions $w(x)$ or $q(x)$ blow up, or $p(x)$ vanishes at some point in $[a,b]$. The Hilbert space associated to (2.27) is $L^2([a,b],w)$ consisting of Lebesgue measurable functions f on $[a,b]$ such that

$$\int_a^b |f(x)|^2 w(x)\,dx < \infty.$$

The **inner product** for this weighted L^2-space with weight w is

$$(f,g) = \int_a^b f(x)\overline{g(x)}w(x)\,dx.$$

If one makes the correct assumptions about p,q,w and if one imposes the correct sort of boundary conditions at a and b, the operator L will be self-adjoint. In particular, p,q,w should be real and w should be positive.

The Green's function $G(\lambda;x,y)$ of formula (2.24) must satisfy

$$-\frac{\partial}{\partial x}\left(p\frac{\partial}{\partial x}G(\lambda;x,y)\right) + (q-\lambda w)G(\lambda;x,y) = \delta(x-y).$$

Pick c in (a,b). We want to choose two linearly independent solutions φ_λ and ψ_λ of $Lu=\lambda u$, as follows. Suppose that φ_λ solves $Lu=\lambda u$ and in addition φ_λ lies in $L^2([a,c],w)$. Let ψ_λ solve $Lu=\lambda u$ and $\psi_\lambda \in L^2([c,b],w)$. Then

$$G(\lambda;x,y) = \frac{1}{c_\lambda}\begin{cases} \varphi_\lambda(x)\psi_\lambda(y), & a<x<y<b \\ \varphi_\lambda(y)\psi_\lambda(x), & a<y<x<b \end{cases} \tag{2.28}$$

and

$$c_\lambda = \begin{vmatrix} \varphi_\lambda(x) & -\psi_\lambda(x) \\ p\varphi'_\lambda(x) & -p\psi'_\lambda(x) \end{vmatrix}.$$

*Example 2.2.1 (**Spectral Measure for the Hankel Transform Using the Kodaira–Titchmarsh Formula**).* Consider the Sturm–Liouville operator given by formula (2.23). Then formula (2.28) becomes

$$G(\lambda;x,y) = \frac{\pi i}{2}\begin{cases} J_\lambda(\lambda^{1/2}x)H_\nu^{(1)}(\lambda^{1/2}y), & 0 < x < y < \infty \\ J_\lambda(\lambda^{1/2}y)H_\nu^{(1)}(\lambda^{1/2}x), & 0 < y < x < \infty. \end{cases}$$

Here J_ν is the Bessel function of the first kind from Exercise 2.2.2 and $H_\nu^{(1)}$ is the Hankel function (or Bessel function of the third kind) defined by

$$H_\nu^1(z) = \frac{J_{-\nu}(z) - \exp(-\nu\pi i)J_\nu(z)}{i\sin(\nu\pi)}$$

when ν is not an integer. Take limits as $\nu \to n$ to define $H_n^{(1)}$ when n is an integer. It follows from Lebedev [401, pp. 112–113], that the determinant $c_\lambda = -2i/\pi$ in this case.

In order to check this formula for $G(\lambda;x,y)$, one needs the following asymptotic properties of the Bessel functions from Lebedev [401, Chap. 5], $(\nu > 0)$:

$$\begin{aligned} &J_\nu(x) \sim \left(\tfrac{x}{2}\right)^\nu \tfrac{1}{\Gamma(1+\nu)}, && \text{as } x \to 0, \\ &J_\nu(x) \sim \sqrt{\tfrac{2}{\pi x}}\cos\left(x - \tfrac{\nu\pi}{2} - \tfrac{\pi}{4}\right), && \text{as } x \to \infty, \\ &H_\nu^{(1)}(x) \sim -\tfrac{i}{\pi}\left(\tfrac{2}{x}\right)^\nu \Gamma(\nu), && \text{as } x \to 0, \\ &H_\nu^{(1)}(x) \sim \sqrt{\tfrac{2}{\pi x}}\exp\left[i\left(x - \tfrac{\nu\pi}{2} - \tfrac{\pi}{4}\right)\right], && \text{as } x \to \infty. \end{aligned} \tag{2.29}$$

Note that if $0 < \nu < 1$, both $J_\nu(\lambda^{1/2}x)$ and $H_\nu^{(1)}(\lambda^{1/2}x)$ are in $L^2([0,1],x)$. One says that 0 is then in the limit-circle case in Weyl's theory (cf. Stakgold [618,619]). If $\nu \geq 1$, then $J_\nu(\lambda^{1/2}x)$ is in $L^2([0,1],x)$ and $H_\nu^{(1)}(\lambda^{1/2}x) \notin L^2([0,1],x)$. One says that 0 is then in the limit-point case. Note that $J_\nu(\lambda^{1/2}x) \notin L^2([1,\infty],x)$ and $H_\nu^{(1)}(\lambda^{1/2}x) \in L^2([1,\infty],x)$ so that infinity is always in the limit-point case.

To compute the jump of $G(\lambda;x,y)$ required by formula (2.26), assume that $x < y$. Then for $\nu > 0$,

$$\begin{aligned} &G(\lambda + i0;x,y) - G(\lambda - i0;x,y) \\ &\quad = \tfrac{i\pi}{2}\left(J_\nu(\lambda^{1/2}x)H_\nu^{(1)}(\lambda^{1/2}y) - J_\nu(-\lambda^{1/2}x)H_\nu^{(1)}(-\lambda^{1/2}y)\right). \end{aligned}$$

So we need the functional equations

$$\begin{aligned}
&J_\nu(-x) = \exp(\nu\pi i)J_\nu(x),\\
&H_\nu^{(1)}(-x) = \exp(-\nu\pi i)\left(H_\nu^{(1)}(x) - 2J_\nu(x)\right).
\end{aligned}$$

These formulas imply

$$G(\lambda + i0; x, y) - G(\lambda - i0; x, y) = i\pi J_\nu(\lambda^{1/2}x)J_\nu(\lambda^{1/2}y).$$

Thus we obtain from (2.24) and (2.26),

$$\frac{\delta(x-y)}{x} = \frac{1}{2}\int_0^\infty J_\nu(\lambda^{1/2}x)J_\nu(\lambda^{1/2}y)\,dy = \int_0^\infty J_\nu(\rho x)J_\nu(\rho y)\,\rho\,d\rho.$$

This is Hankel's integral theorem.

For higher-rank symmetric spaces, the method just used to compute the spectral measure proves rather unwieldy because the Green's functions are much more complicated. Thus we will formulate another method in Chap. 3 and Volume II [667]. For this method, one needs only the asymptotics and functional equations of the eigenfunctions appearing in the spectral expansion of the operator L, rather than the expression for the resolvent kernel, involving other eigenfunctions as well. The method is due to Harish-Chandra and it would be very interesting to derive Harish-Chandra's Plancherel measure from the spectral theorem in a similar way to that which gives formula (2.26).

One can also ask for the spectral decomposition of the operator in (2.23) on the finite interval $(0,a)$. The problem still has a singularity at 0. But on $(0,a)$, the spectral resolution of the operator L in (2.23) is given by a series rather than an integral, when $\nu > -1$:

$$f(x) = \sum_{m=1}^{\infty} c_m J_\nu(\alpha_{\nu m}x/a), \quad 0 \le x \le a, \quad \nu > -1, \tag{2.30}$$

where $\{\alpha_{\nu m}\}_{m\ge 1}$ is the set of all positive roots of $J_\nu(\alpha_{\nu m}) = 0$ and

$$c_m = 2a^{-2}[J_{\nu+1}(\alpha_{\nu m})]^{-2}\int_0^a f(r)J_\nu(\alpha_{\nu m}r/a)r\,dr.$$

The **Fourier–Bessel series expansion** (2.30) is derived in Titchmarsh [679, Vol. 1, pp. 81–86]. See also Stakgold [619, pp. 305–308, 313–315]. It is quite interesting to let a approach infinity in (2.30) and watch (2.30) approach (2.22) (cf. Morse and Feshbach [479, Vol. I, pp. 762–766]).

Exercise 2.2.8 (The Kontorovich–Lebedev Transform). Consider Bessel's equation with the role of the parameters interchanged:

$$-(xw')' + \mu x w - \lambda w/x = 0, \quad 0 < x < \infty.$$

Here $\mu > 0$ and λ is the eigenvalue. Show that the Green's function is

$$G(\lambda;x,y) = \begin{cases} I_{-i\sqrt{\lambda}}(\mu^{1/2}x)K_{-i\sqrt{\lambda}}(\mu^{1/2}y), & 0<x<y<\infty, \\ I_{-i\sqrt{\lambda}}(\mu^{1/2}y)K_{-i\sqrt{\lambda}}(\mu^{1/2}x), & 0<y<x<\infty, \end{cases}$$

where I and K are the **Bessel functions of imaginary argument**. The definitions are

$$\begin{aligned} I_\nu(z) &= \exp(-\nu\pi i/2)J_\nu(z\exp(\pi i/2)), \quad -\pi < \arg z < \pi/2, \\ K_\nu(z) &= \frac{\pi i}{2}\exp(\nu\pi i/2)H_\nu^{(1)}(z\exp(\pi i/2)), \quad -\pi < \arg z < \pi/2. \end{aligned}$$

See Chap. 3 and Lebedev [401] for more information on these functions.

Show that $K_{-\nu}(x) = K_\nu(x)$ and $I_{-\nu}(x) - I_\nu(x) = 2\sin(\nu\pi)K_\nu(x)/\pi$. Then prove the **Kontorovich–Lebedev inversion formula**:

$$x\delta(x-y) = \frac{2}{\pi^2}\int_0^\infty K_{i\nu}(\sqrt{\mu}x)K_{i\nu}(\sqrt{\mu}y)\ \nu\sinh(\pi\nu)\ d\nu.$$

This will be a central result in Sect. 3.1.

2.2.2 *CAT Scanners and the Radon Transform*

Modern X-ray scanners can reproduce the tissue density function $f(x),\ x \in \mathbb{R}^2$, for a plane slice of a person's head, for example. They operate by inverting the **Radon transform**, defined for $k \in \mathbb{R},\ u \in S^1$, by

$$Rf(k,u) = \int_{\substack{{}^t xu=k \\ x\in\mathbb{R}^2}} f(x)\ ds, \tag{2.31}$$

where the integral is over a line in the plane and ds is the element of arc length. The **inversion formula** for this transform goes back to Radon [528] in 1917. It says that

$$f(x) = -\frac{1}{\pi}\int_{q>0}\frac{1}{q}d\left(\frac{1}{2\pi}\int_{u\varepsilon S^1} Rf(q+{}^t xu,u)\ du\right). \tag{2.32}$$

Here the integral over q in $\mathbb{R}^+$ can be viewed as a Stieltjes integral or as a Cauchy principal value integral, assuming that f is continuous with compact support.

References for the Radon transform include Bracewell [62], Dym and McKean [147], Gelfand, Graev, and Vilenkin [204], Herman [290], Helgason [279], Louis [431, 432], Ludwig [435], Elena Prestini [522], Quinto [524], and Shepp and Kruskal [588]. In fact, Funk [187] proved the analogue of (2.32) for S^2 rather than

$\mathbb{S}^1$ one year earlier than Radon proved his result. Helgason [279] shows that a vast generalization of this theory is possible, viewing the Radon inversion formula as involving two dual integrations—one over the points in a hyperplane and the other over the hyperplanes through a point. The Radon transform appears in so many applications to signal processing that Matlab's Signal Processing Toolbox has a Radon transform command. We look at finite analogues of the Radon transform in [668, pp. 71–72]

The next sequence of exercises presents a derivation of Radon's inversion formula.

Exercise 2.2.9. (a) Suppose that $f : \mathbb{R}^2 \to \mathbb{C}$ is a Schwartz function. Show that the Fourier inversion formula can be written in the form

$$f(x) = \int_{u\in S^1} Ef({}^t ux, u)\, du,$$

where

$$Ef(t,u) = \frac{1}{2}\int_{r\in\mathbb{R}}\int_{k\in\mathbb{R}} |r| Rf(k,u) \exp[2\pi i r(t-k)]\, dk\, dr.$$

In the first formula du denotes the angle measure on the unit circle S^1. The Radon transform $Rf(k,u)$ is defined by (2.31).

(b) Let $\text{abs}(r) = |r|,\ r \in \mathbb{R}$. Show that $\text{abs}(r)$ is not the Fourier transform of a Lebesgue integrable function.

Hint. *Recall the Riemann–Lebesgue lemma. Note that if* $\widehat{\text{abs}(r)}$ *were a function, then you could write:*

$$Ef(t,u) = \frac{1}{2}\int_{k\in\mathbb{R}} \widehat{\text{abs}(t-k)} Rf(k,u)\, dk.$$

Exercise 2.2.10 (Derivatives and Fourier Transforms of Some Distributions). Define the Cauchy principal value integral by

$$PV\left(\int f(x)\, dx\right) = \lim_{\varepsilon\to 0}\left(\int_{|x|>\varepsilon} f(x)\, dx\right).$$

Define

$$(x^{-1},\varphi) = PV\int x^{-1}\varphi(x)\, dx$$

and

$$(x^{-2},\varphi) = PV\int x^{-2}(\varphi(x)-\varphi(0))\, dx$$

for test functions φ as in Sect. 1.1. Prove that in the sense of distributions (see Sect. 1.1) the following formulas are valid:

(a) $x^{-1} = (\log|x|)'$.
(b) $x^{-2} = -(x^{-1})'$.
(c) $\widehat{x^{-1}} = -i\pi \operatorname{sgn}(x)$, where $\operatorname{sgn}(x) = \operatorname{abs}(x)/x$.
(d) $\widehat{\operatorname{abs}(x)} = -(2\pi^2)^{-1}x^{-2}$.

Hint. *See Vladimirov [706, pp. 75, 86, 134], or Bracewell [61, p. 130].*

Exercise 2.2.11. Derive Radon's inversion formula from Exercises 2.2.9 and 2.2.10, using properties of Fourier transforms of distributions.

Note. The **Hilbert transform** of a function is

$$Hf(x) = -\frac{1}{\pi}\int_{-\infty}^{+\infty} \frac{f(t)}{(x-t)}\,dt = -\frac{1}{\pi} f * (x^{-1}).$$

Thus part (c) of Exercise 2.2.10 implies that

$$\widehat{Hf(x)} = i \operatorname{sgn}(x)\hat{f}.$$

Because Radon's inversion formula (2.32) contains a derivative, it does not appear to be as useful for numerical calculation as the formula of Exercise 2.2.9. So those that have computed these transforms in practice have used approximations to $\widehat{\operatorname{abs}(x)}$ (cf. Herman [290, pp. 19ff], and Shepp and Kruskal [588]).

Exercise 2.2.12. Compute the Fourier transform of the following approximation to $\operatorname{abs}(x)$ for $x \in \mathbb{R}^2$:

$$f(x) = \begin{cases} |x|, & \text{if } |x| < A, \\ 0, & \text{otherwise.} \end{cases}$$

Compare the decay at infinity with that of $|x|^{-2}$ from part (d) of Exercise 2.2.10.

Note. You can write the Fourier transform in the plane as a Bessel transform, much as we did in Theorem 2.2.1 for the Fourier transform in 3-space. See Exercise 2.2.5.

Some History. In 1956, R. Bracewell used a method analogous to the Radon transform to study solar radiation. In the 1960s and 1970s CT scanners for medicine were developed independently by A. M. Cormack and G. N. Hounsfield. They were not based on the Radon transform. The Hounsfield algorithm was used in the first commercial CAT scanner made by EMI Central Research Labs. in the UK. The first patient brain scan was done in 1971. A. M. Cormack and G. N. Hounsfield shared the Nobel Prize in Medicine in 1979.

The four illustrations in Figs. 2.6 and 2.7 from the 1978 paper of Shepp and Kruskal [588] show a mathematical phantom on the bottom in Fig. 2.6 representing

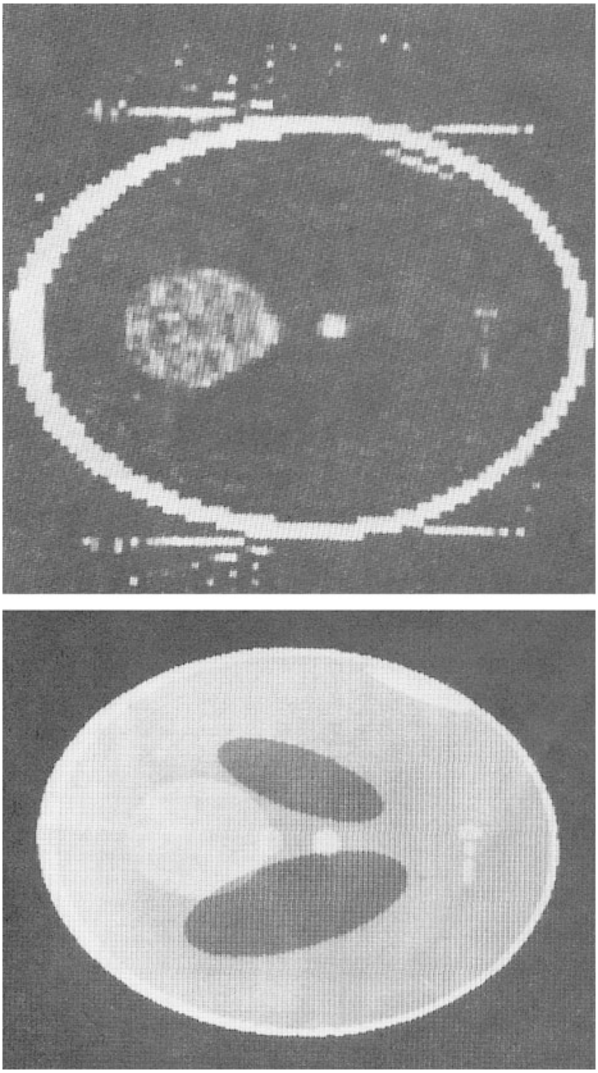


Fig. 2.6 On the bottom is a simulation of a human head using 11 ellipses. On the *top* is a reconstruction using the algorithm embodied in the first commercial machine (EMI Ltd.) from 180×160 strip projection data obtained by exact calculation from the image on the *bottom*. (From Shepp and Kruskal [588, pp. 422–423]. Reprinted by permission of *American Mathematics Monthly*)

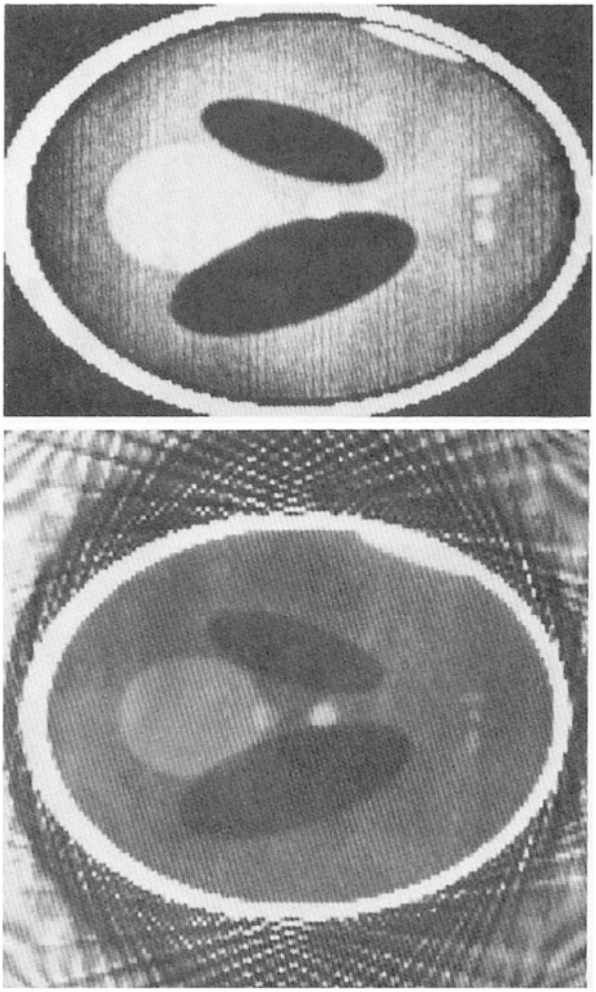


Fig. 2.7 On the bottom is a reconstruction from the *bottom image* in Fig. 2.6 using the Fourier based algorithm of Shepp. On the *top* is a reconstruction using the algorithm used in the 1970s EMI machine from 180 × 239 strip projection data obtained by exact calculation from the *bottom image* in Fig. 2.6. (From Shepp and Kruskal [588, p. 424]. Reprinted by permission of *American Mathematics Monthly*

a slice of a human head and reconstructions of this phantom by three different algorithms. The Hounsfield algorithm was used to produce the top part of Fig. 2.6. The bottom part of Fig. 2.7 was produced with a Fourier-based algorithm of Shepp. The top part of Fig. 2.7 was produced with an algorithm used by EMI Ltd. in the 1970s. Clearly progress was made.

A description of the progress made between the 1970s and 1980s is contained in the following quotation from *Science 214* (1981), p. 1327:

> When CT scanners first became commercially available about 8 years ago it took 5 minutes to scan a patient's head and 5 minutes for each computerized reconstruction of an image from the *x*-ray data. Now, because of advances in the design of the scanners and in computer technology, the newest machines can scan a head just 10 seconds and can reconstruct an image virtually instantaneously. According to Jay Thomas Payne of Abbott Northwestern Hospital in Minneapolis, the Mayo Clinic's first CT scanner, which is only 5 years old, has been relegated to the clinic's historical museum.

Of course, despite the progress in speed and accuracy of CT scanners, they still expose people to radiation. In a *New York Times* article from August 21, 2012, Jane E. Brody wrote the following.

> But it [radiation] also has a potentially serious medical downside: the ability to damage DNA and, 10 to 20 years later, to cause cancer. CT scans alone, which deliver 100 to 500 times the radiation associated with an ordinary X-ray and now provide three-fourths of Americans' radiation exposure, are believed to account for 1.5 percent of all cancers that occur in the United States.

Thus nuclear magnetic resonance tomography (NMR alias MRI) may be destined to be the tomography of the future. It uses magnetic fields rather than *x*-rays and thus is *presumably* less damaging to the body. Louis [431,432] considers applications of the three-dimensional Radon transform to NMR tomography. Spherical harmonics are used to improve the algorithm by studying the kernel (i.e., inverse image of 0) of the transform (ghosts). In 2003, the Nobel Prize for Physiology or Medicine went to Paul Lauterbur and Sir Peter Mansfields for their work on MRI. See Elena Prestini [522], Chap. 8, for more details on the subject. She begins the chapter with a quotation from a *New York Post* article from 1939 about a talk by I. I. Rabi explaining why NMR scanning works. The quote is: "We are all radio stations."

There are many other applications of the Radon transform; for example, in radio astronomy (see Bracewell's article in Herman [290, pp. 81–104]). And there are applications to partial differential equations; e.g., to find solutions of the wave equation (see Dym and McKean [147, pp. 137–139], and Helgason [279].

This concludes our discussion of harmonic analysis on the sphere. It would be possible to consider spherical analogues of many more results from Chap. 1. We shall leave this to the interested reader. For example, the group of isometries of the sphere is just $O(3)$. A discontinuous subgroup $\Gamma \subset O(3)$ has the property that any domain in the sphere can contain only finitely many points equivalent under Γ to any given point. There are very few such discontinuous Γ for the sphere or elliptic plane. They correspond to the regular polyhedra (see Hilbert and Cohn-Vossen [297, p. 242]). These are the analogues of the space groups considered in Sect. 1.4. It is also possible to discuss the analogue of the Poisson summation formula for $\Gamma\backslash S^2 \cong \Gamma\backslash G/K,\ G = SO(3),\ K = SO(2)$. We shall not do this here, since our main interest is the Selberg trace formula for noncompact fundamental domains (see Chap. 3). In fact, we never examined the Selberg trace formula when G is the Euclidean group either (but see Hejhal [265]).

There is work on fast computation of Fourier transforms on the sphere by Driscoll and Healy [140], for example. There are also finite analogues of Radon transforms (see [668]).

Chapter 3
The Poincaré Upper Half-Plane

> But why then, this mystical set-up of putting the definition before the proof?
>
> —From Lakatos [384, p. 154].

3.1 Hyperbolic Geometry

> … la géométrie non-Euclidienne… est la clef véritable du problème qui nous occupe.[1]
>
> —From a letter of Poincaré to Klein in 1881 appearing in *Acta Math., 39* (1923), p. 100.

3.1.1 The Basics: Arc Length, Area, Laplacian

In Chap. 2, we considered a model for elliptic geometry, in which any two geodesics intersect, so that there are no parallels. Now we want to investigate a model for hyperbolic geometry, in which there are infinitely many geodesics through a given point which are parallel to a given geodesic. This sort of geometry was discovered by Bolyai, Gauss, and Lobatchevsky in the 1820s. However, Gauss never published his results, perhaps because the idea was controversial. In fact, Gauss embittered Bolyai by claiming precedence in a letter to Bolyai's father (see Gauss [199, Vol. 8, pp. 220–225]). The subject of non-Euclidean geometry was still controversial when Lewis Carroll "repudiated hyperbolic geometry in 1888 as being too fanciful" (see the article of Coxeter in *COSRIMS* [110, p. 55]). Models for hyperbolic geometry were obtained first by Liouville, then by Beltrami in 1868, and by Klein in 1870. Poincaré rediscovered the Liouville–Beltrami upper half-plane model in 1882 and

---

[1] Non-Euclidean geometry...is the real key to the problem that concerns us.

A. Terras, *Harmonic Analysis on Symmetric Spaces—Euclidean Space, the Sphere, and the Poincaré Upper Half-Plane*, DOI 10.1007/978-1-4614-7972-7_3,
© Springer Science+Business Media New York 2013

this space is usually called the Poincaré upper half-plane, though some call it the Lobatchevsky upper half-plane (but see Milnor [469]). Poincaré [517] also considered discontinuous groups of transformations of the hyperbolic upper half-plane as well as the functions left invariant by these groups and we intend to do the same. The geometric foundations for such work were laid by Gauss in 1827 (see Gauss [199, Vol. 4, pp. 217–258]) and by Riemann in 1854 (see Riemann [542, pp. 272–287]). These important papers of Gauss and Riemann are discussed from a modern perspective in Spivak [615, Vol. II].

Further progress was made from the 1880s to the 1930s with the investigation of Lie groups and symmetric spaces by Lie, Cartan, and others. Ultimately, non-Euclidean geometry has become about as controversial as $\sqrt{-1} = i$. Like complex analysis, analysis on Lie groups and symmetric spaces such as the Poincaré upper half-plane has become an indispensable tool for modern work in physics and engineering. We will see many examples, including work on the design of microwave transmission lines and the computation of the electrostatic field due to a thin charged conductor in the shape of the surface of two intersecting spheres. There are many applications to quantum physics in Gutzwiller [239–242], and Hurt [308,309]. Some references for this section are Auslander [17], Beardon [31], Buser [74], Dym and McKean [147], Hilbert and Cohn-Vossen [297], Svetlana Katok [342], Maass [437], and Siegel [596].

We shall see that the Poincaré upper half-plane is the symmetric space of the Lie group $SL(2,\mathbb{R})$, the **special linear group** of all 2×2 real matrices with determinant 1. This group has made many appearances in number-theoretic investigations, going back to work of Gauss and others on quadratic forms with integer coefficients. The most recent appearance is in the proof of Fermat's conjecture by A. Wiles with the help of R. Taylor. We will say a little more about this in Sect. 3.4.

Now $SL(2,\mathbb{R})$ is isomorphic to the group $SU(1,1)$ of 2×2 complex matrices of determinant one which preserve the hermitian form $-z_1\overline{z_1} + z_2\overline{z_2}$. This means that the corresponding symmetric space can also be viewed as the unit disc (see Sect. 3.1.2 with its applications to electrical engineering). The group $SL(2,\mathbb{R})$ is locally isomorphic to the Lorentz-type group $SO(2,1)$ of real 3×3 matrices of determinant 1 preserving the quadratic form $x_1^2 + x_2^2 - x_3^2$, a group whose analogue $SO(3,1)$ is quite important in physics, since it leaves Maxwell's equations invariant (see Exercise 3.1.13).

The **Poincaré or hyperbolic upper half-plane** H is defined by

$$H = \{x + iy \mid x, y \in \mathbb{R}, y > 0\}, \quad \text{where} \quad i = \sqrt{-1}, \tag{3.1}$$

using a new notion of **arc length** given by

$$ds^2 = y^{-2}(dx^2 + dy^2). \tag{3.2}$$

This arc length is not "too fanciful," because it is invariant under the **action** of g in $SL(2,\mathbb{R})$ on z in H defined by fractional linear transformation:

$$gz = g(z) = (az+b)/(cz+d), \tag{3.3}$$
$$\text{if} \quad g = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \text{ with } a,b,c,d \in \mathbb{R}, \quad ad-bc = 1.$$

More information on fractional linear transformations can be found in Ahlfors [3], Siegel [596, Chap. 3], and almost any book on complex analysis.

Exercise 3.1.1. Show that the map $z \mapsto g(z)$ defined by Eq. (3.3) gives a conformal or angle-preserving mapping of H onto H.

Hint. *Note that* $\operatorname{Im}(g(z)) = y|cz+d|^{-2}$, *if* $z = x+iy$, *and* $g'(z) = (cz+d)^{-2}$.

Exercise 3.1.2 (Group Invariance of the Riemannian Metric on H). Show that the Poincaré arc length element (3.2) is invariant under the action (3.3) of g in $SL(2,\mathbb{R})$.

Hint. *If* $w = f(z)$ *is a holomorphic function and* $w = u+iv, z = x+iy$, *with* $u,v,x,y \in \mathbb{R}$, *then the Jacobian matrix of the change of variables from* z *to* w *is*

$$J = \begin{pmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \end{pmatrix} = \begin{pmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ -\frac{\partial u}{\partial y} & \frac{\partial u}{\partial x} \end{pmatrix},$$

by the Cauchy–Riemann equations. Now

$$\begin{pmatrix} du \\ dv \end{pmatrix} = J \begin{pmatrix} dx \\ dy \end{pmatrix},$$

and the determinant of J *is* $|f'(z)|^2$. *In our case* $f(z) = (az+b)/(cz+d)$ *with* $a,b,c,d \in \mathbb{R}$ *and* $ad-bc = 1$. *Thus, you easily compute that*

$$v = y|cz+d|^{-2} \text{ and } f'(z) = (cz+d)^{-2}.$$

The **geodesics**, or curves minimizing the Poincaré arc length in H, are straight lines or semi-circles orthogonal to the real axis. To prove this, we imitate the argument that works for Euclidean space and the sphere. It suffices to show that the positive y-axis is the curve minimizing distance and passing through i and $y_0 i$, where $y_0 > 0$, for we can find an element g of $SL(2,\mathbb{R})$ that moves any two given points in H to i and iy_0, for some $y_0 > 0$, by Exercise 3.1.3. Now let

$$z(t) = x(t) + iy(t), \quad a \leq t \leq b,$$

denote any curve in H with $z(a) = i$ and $z(b) = iy_0$. Then the Poincaré length of this curve is (for $y_0 > 1$)

$$\int_a^b y^{-1} \left(x'^{\,2} + y'^{\,2}\right)^{\frac{1}{2}} dt \geq \int_1^{y_0} y^{-1} dy = \log y_0.$$

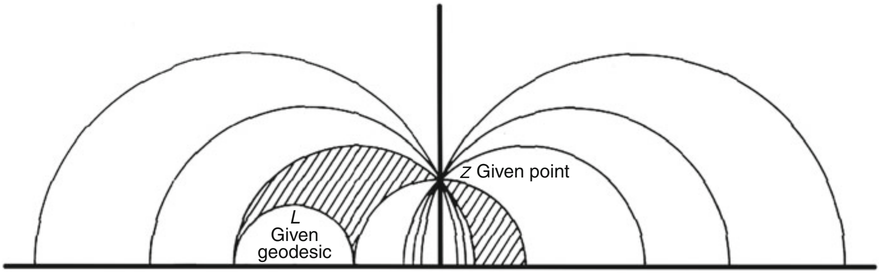


Fig. 3.1 The failure of Euclid's fifth postulate. All geodesics through z outside the shaded angle fail to meet L

And $|\log y_0|$ is the Poincaré length of the segment of the y-axis joining i and iy_0. We will see that a similar argument works in the higher-dimensional symmetric spaces to be considered in Volume II [667].

Exercise 3.1.3. Show that given p,q in H, there is a matrix $g \in SL(2,\mathbb{R})$ such that $g(p) = i$ and $g(q) = iy_0$ for some $y_0 > 0$.

Hint. *First, move* p *to* i. *Then note that* $k \in SO(2) \subset SL(2,\mathbb{R})$ *implies* $k(i) = i$.

We can use the Poincaré arc length to define the **non-Euclidean distance** between two points in H. This distance is the Poincaré length of the unique geodesic segment connecting the two points. The geometry obtained by considering geodesics to be the straight lines is called **hyperbolic geometry**. It is similar to Euclidean geometry in many respects. For example, two points in H determine a unique geodesic passing through them. And the non-Euclidean distance makes H a metric space. So, for example, the distance on H satisfies the triangle inequality, meaning that the length of one side of a non-Euclidean triangle is less than or equal to the sum of the lengths of the other two sides.

Exercise 3.1.4. Show that two points in H determine a unique geodesic passing through them. Then show that the non-Euclidean distance satisfies the triangle inequality.

Exercise 3.1.5. We can use the Poincaré metric, which is really an inner product on the tangent space to H at a point, to define the angle between tangent vectors or curves in H. Show that this notion of angle is the same as the Euclidean angle measure.

But the hyperbolic geometry does not satisfy Euclid's fifth postulate. For given a point z in H not on a geodesic L, there are infinitely many geodesics through z which do not meet L (see Fig. 3.1, in which all geodesics through z outside the shaded region fail to meet L).

The group of all isometries of H is generated by the fractional linear transformations (3.3) and $z \mapsto -\overline{z}$. Here isometry means Poincaré-distance preserving. See Exercise 3.1.12 for a proof.

We can now use the standard formulas from Riemannian geometry to write down the $SL(2,\mathbb{R})$-invariant area element and Laplacian (see formulas (2.2)–(2.6) and Auslander [17]). The $SL(2,\mathbb{R})$-invariant **area element** on H is

$$d\mu = y^{-2} dx\, dy. \tag{3.4}$$

The **Laplace operator** on H is

$$\Delta_H = \Delta = y^2 \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \right). \tag{3.5}$$

Exercise 3.1.6 (SL(2,$\mathbb{R}$)-Invariance of the Area Element and the Laplacian on H). Show that $d\mu$ in Eq. (3.4) and Δ_H in Eq. (3.5) are $SL(2,\mathbb{R})$ invariant.

Hint. *Recall Exercise 3.1.2 and formulas (2.2)–(2.6).*

Exercise 3.1.7. Show that the non-Euclidean area of a hyperbolic triangle is π minus the sum of the angles.

Hint. *You can prove this directly or note that this is a special case of the Gauss–Bonnet formula, since the Gaussian curvature of H is* -1 *(see Singer and Thorpe [604, pp. 175, 191–192], or Auslander [17, pp. 265, 268]).*

It is now possible to explain the term ***hyperbolic*** in at least two ways. The term is due to Klein and comes from the Greek word *hyperballein* meaning to throw beyond. The first justification is that if two geodesic rays R_1 and R_2 emanate from the ends of a geodesic segment perpendicular to them both, then the non-Euclidean distance between R_1 and R_2 will increase. The second explanation comes from the fact that the Gaussian curvature of H is -1 and thus H resembles a hyperboloid of one sheet or a hyperbolic paraboloid (e.g., $z = x^2 - y^2$ at the origin). References for this are Auslander [17], Flanders [180], Guggenheimer [235], and Hilbert and Cohn-Vossen [297]. We justified the term *elliptic* similarly in Chap. 2.

There are other useful realizations of H. The most abstract of them is given in the following exercise. Compare it with Exercise 2.1.1 of Sect. 2.1.

Exercise 3.1.8 (H as a Quotient or Homogeneous Space). Show that H can be identified with G/K where $G = SL(2,\mathbb{R})$ and $K = SO(2)$.

Hint. *Map* G/K *one-to-one onto H by sending the coset* $gK, g \in G$*, to* $g(i) =$ *the image of* $i = \sqrt{-1}$ *under the fractional linear transformation (3.3) corresponding to g.*

For connecting Chap. 3 and Volume II, [667] the most useful realization of H is as **the space of positive definite binary quadratic forms of determinant 1** or what amounts to the same thing:

$$\mathcal{SP}_2 = \left\{ Y \in \mathbb{R}^{2\times 2} \;\middle|\; Y \text{ positive definite symmetric of determinant } 1 \right\}. \tag{3.6}$$

A 2×2 symmetric matrix Y with real coefficients is **positive** if $Y[x] = {}^t xYx > 0$ for every nonzero column vector x in $\mathbb{R}^2$. See Volume II [667] for more information about positive matrices. Note that the **action** of g in $SL(2,\mathbb{R})$ on Y in $\mathcal{SP}_2$ is given by

$$Y \to Y[g] = {}^t gYg. \tag{3.7}$$

We shall use left G-actions here in Chap. 3 and right G-actions in Volume II [667]. Hopefully this will not cause too much confusion.

Exercise 3.1.9 (H as a Space of Positive Quadratic Forms of Determinant 1).

(a) Show that the following maps are identifications and preserve the group action of $SL(2,\mathbb{R})$ on the three homogeneous spaces:

$$K\backslash G \to \mathcal{SP}_2 \to H$$
$$Kg \mapsto {}^t gg = P \mapsto z \in H \text{ with } P\begin{bmatrix} z \\ 1 \end{bmatrix} = 0,$$

where $P[A] = {}^t APA$ for any suitable matrix or vector A, as in formula (3.7). Here $K\backslash G$ is used to denote the homogeneous space of right cosets Kg, $g \in G = SL(2,\mathbb{R})$, $K = SO(2)$. Note that you can write P in $\mathcal{SP}_2$ as follows:

$$P = \begin{pmatrix} y^{-1} & 0 \\ 0 & y \end{pmatrix} \begin{bmatrix} 1 & -x \\ 0 & 1 \end{bmatrix} = \begin{pmatrix} y^{-1} & -x/y \\ -x/y & (x^2+y^2)/y \end{pmatrix},$$
$$\text{with} \quad x, y \in \mathbb{R},\ y > 0.$$

This is clearly possible. Then $P\begin{bmatrix} z \\ 1 \end{bmatrix} = 0$ for $z \in H$ implies $z = x + iy$.

(b) Show that a geodesic-reversing isometry of $\mathcal{SP}_2$ at the identity I is $Y \mapsto Y^{-1}$. The corresponding mapping in H is $z \mapsto -z^{-1}$.

Given $P \in \mathcal{SP}_2$, the spectral theorem for self-adjoint operators on finite-dimensional Hilbert space says there is a matrix k in $SO(2)$ such that $P[k] = {}^t kPk$ is diagonal with first diagonal entry $t > 0$ and second entry $1/t$. We can think of k and t as **polar coordinates** for P. If you want to use geodesic polar coordinates, you would use k and $|\log t|$. This amounts to defining **geodesic polar coordinates** (r,u) of $z = x + iy$ in H by

$$z = \begin{pmatrix} \cos u & \sin u \\ -\sin u & \cos u \end{pmatrix} \begin{pmatrix} \exp(-r/2) & 0 \\ 0 & \exp(r/2) \end{pmatrix} (i),$$
$$x = y \sinh r\ \sin(2u),\ y = (\cosh r + \cos(2u)\ \sinh r)^{-1}. \tag{3.8}$$

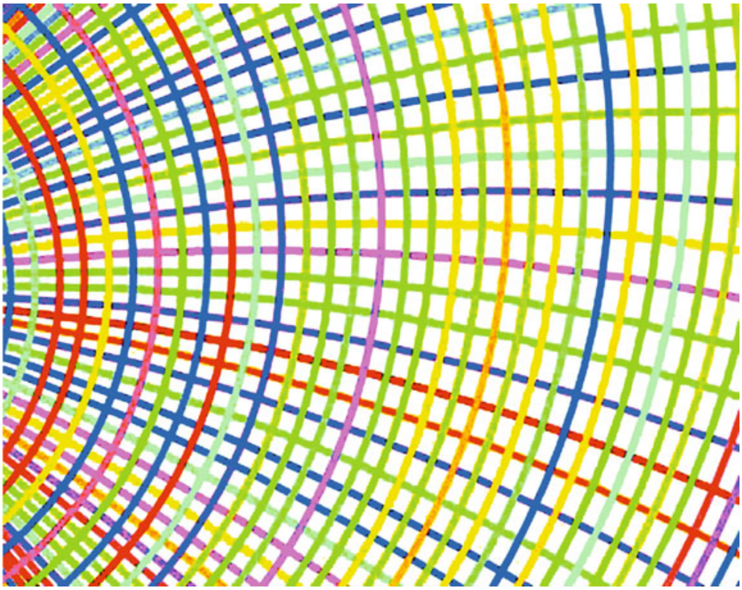


Fig. 3.2 Coordinate grid for geodesic polar coordinates

As u runs from 0 to 2π and r from 0 to infinity, the upper half-plane is covered once. To see this, note that the eigenvalues t of the positive definite matrix $P \in \mathcal{SP}_2$ corresponding to z as in Exercise 3.1.9 are uniquely determined up to order. Formula (3.8) fixes that order by requiring that the first eigenvalue be the smaller one.

Let K_P denote the subgroup of $k \in SO(2)$ preserving a diagonal matrix, $P \in \mathcal{SP}_2$, under the action $P[k] = {}^t kPk$. This subgroup K_P has order 2 if P is not a scalar times the identity matrix. If $P \neq cI$, for $c > 0$, the subgroup K_P does not affect elements of H at all.

Exercise 3.1.10. Show that in formula (3.8), r is the Poincaré distance between z and $i = \sqrt{-1}$. Then show that the orthogonal grid in H created by the curves $u =$ constant and $r =$ constant consists of circles and semi-circles, such as those pictured in Fig. 3.2. In particular,

$$\{k(e^{-r}i) \mid r > 0\} = \text{semicircle through } i \text{ orthogonal to the} x\text{−axis},$$

$$\{k(e^{-r}i) \mid k \in SO(2)\} = \text{a circle with center} \quad i\cosh r \quad \text{and} \quad \text{radius} \quad \sinh r.$$

Exercise 3.1.11 (Changing Variables in Geodesic Polar Coordinates). Show that in geodesic polar coordinates (r,u) as in Eq. (3.8), we have

$$\begin{aligned}
ds^2 &= dr^2 + (\sinh r)^2 du^2,\\
d\mu &= y^{-2}dx\,dy = \sinh r\,dr\,du,\\
\Delta &= \tfrac{1}{\sinh r}\tfrac{\partial}{\partial r}\left(\sinh r\tfrac{\partial}{\partial r}\right) + \tfrac{1}{(\sinh r)^2}\tfrac{\partial^2}{\partial u^2}.
\end{aligned}$$

Use formulas (2.2)–(2.6).

Another realization of hyperbolic geometry is the **unit disc**:

$$U = \{\, z \in \mathbb{C} \mid |z| \leq 1\}. \tag{3.9}$$

The **Cayley transform**

$$z \mapsto w = (z-i)(z+i)^{-1} \tag{3.10}$$

maps H conformally onto U. See Siegel [596, Chap. 3], Maass [437, Chap. 1], and Helgason [276, pp. 4–7], for more information on this model. The group of orientation-preserving isometries of U is $SU(1,1)$ which is the group of matrices $g \in \mathbb{C}^{2\times 2}$ such that

$${}^t\overline{g}\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}g = \begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}.$$

Such g have the form

$$g = \begin{pmatrix}a & b\\ \overline{b} & \overline{a}\end{pmatrix}, \quad \text{with } a,b \in \mathbb{C} \text{ and } |a|^2 - |b|^2 = 1.$$

This version of hyperbolic geometry is much used by electrical engineers, as we shall see.

Exercise 3.1.12. Use Schwarz's lemma from complex analysis (see Ahlfors [3, p. 135]) to show that the set of all orientation-preserving isometries of the unit disc U is the group $SU(1,1)$. Show also that $SU(1,1)$, together with the isometry given by $z \mapsto \overline{z}$, generates all the isometries of U. Then use the Cayley transform from formula (3.10) to translate this result to H. Finally, characterize the geodesics in U, using the appropriate $SU(1,1)$-invariant arc length.

3.1.2 Microwave Engineering: The Smith Chart

Microwave engineers deal with electromagnetic waves at high frequencies and short wavelengths from 1 m to 1 mm (see the chart of the electromagnetic spectrum appearing in Fig. 1.8). Different books have slightly different numbers for this definition of microwaves. World War II saw the first use of microwaves in order to detect planes and ships via radar. Now there are numerous applications to such diverse areas as astronomy, communications, cooking, location of speeders on the highways, nuclear physics, GPS, satellite TV, weather prediction, wireless local area computer networks, motion detectors, cell phones. Non-Euclidean geometry comes into the design of microwave transmission lines because the basic quantities of interest are related by fractional linear maps (or higher-dimensional analogues). Here we consider only a very simple example in which power transfer is increased on a lossless transmission line by connecting short-circuited stubs to the line. Further information on microwave theory can be found in the books of Altman [5], Baden Fuller [20], Collin [99], Helszajn [285], Hewlett-Packard [293], Lance [385], Magnusson [447], Pozar [521], and Staniforth [620], for example. See Border [50] and Helton [286–288], for a general treatment of power transfer problems involving more complicated circuits with waves of varying frequency.

Here we shall demonstrate that we are not engineers by using $i=\sqrt{-1}$ and not j. We shall also write $\overline{z}$ for the complex conjugate of z and not z^*, as the engineers do.

At short wavelengths, electronic circuits behave differently than they do at long wavelengths which are larger than the dimensions of the circuit under consideration. For example. the circuit parameters vary with position and radiation becomes important. If a voltage wave $V^+\exp(-i\beta d)$ with current $I^+\exp(-i\beta d)$ arrives at the end of a lossless transmission line, a reflected voltage wave $V^-\exp(i\beta d)$ with current $-I^-\exp(i\beta d)$ is produced. Here the phase constant is $\beta = 2\pi/\lambda$, if $\lambda =$ wavelength, and d measures distance along the line. If the load is at $d = 0$, $V_L = V^+ + V^-$, $I_L = I^+ - I^-$. The **load impedance** is $Z_L = V_L/I_L$ and the **characteristic impedance** is $Z_c = V^+/I^+ = V^-/I^-$. Define the **voltage reflection coefficient** $\Gamma = V^-/V^+$ and the **normalized load impedance** $z = Z_L/Z_c$. It follows that

$$\Gamma = (z-1)(z+1)^{-1} \text{ and } z = (1+\Gamma)(1-\Gamma)^{-1}. \tag{3.11}$$

The load is said to be **matched** with the line when $\Gamma = 0$. This is the condition for maximizing power transfer because the **power flow**

$$P = V^+\overline{I}^+ - V^-\overline{I}^- = |V^+|^2\{1-|\Gamma|^2\}/\overline{Z}_c.$$

These definitions extend to give the reflection coefficient at position $-d$:

$$\Gamma(d) = [V^-\exp(-i\beta d)][V^+\exp(i\beta d)]^{-1} = \Gamma_L\exp(-2i\beta d). \tag{3.12}$$

Then the normalized input impedance (seen looking toward the load) at position $-d$ is

$$z_{in} = (1+\Gamma(d))(1-\Gamma(d))^{-1}. \tag{3.13}$$

For lossy lines, $\Gamma(d) = \Gamma_L\exp(-2i\beta d - 2\alpha d)$.

Now $|\Gamma| \leq 1$ and the mapping Eq. (3.13) from z_{in} to Γ is a slight variation of the Cayley transform just discussed. Write $z_{\text{in}} = R + iX$, where $R =$ input resistance and $X =$ input reactance (both normalized). Plotting the image of the lines $R =$ constant and $X =$ constant in the Γ-plane gives the picture in Fig. 3.3. This sort of graph is called a "Smith Chart" because it was proposed by P. H. Smith at Bell Telephone Labs in 1939. However, Matsumoto [457, p. 41], reports that Mitzuhashi had discussed such a chart in 1937. A copy of the actual graph paper (once available in any university bookstore and now available in Matlab and Mathematica) is given in Fig. 3.4. Note that a movement a distance d along the transmission line changes Γ by $\exp(-2\beta i d)$; i.e., the point on the Smith Chart rotates by an angle $2\beta d$.

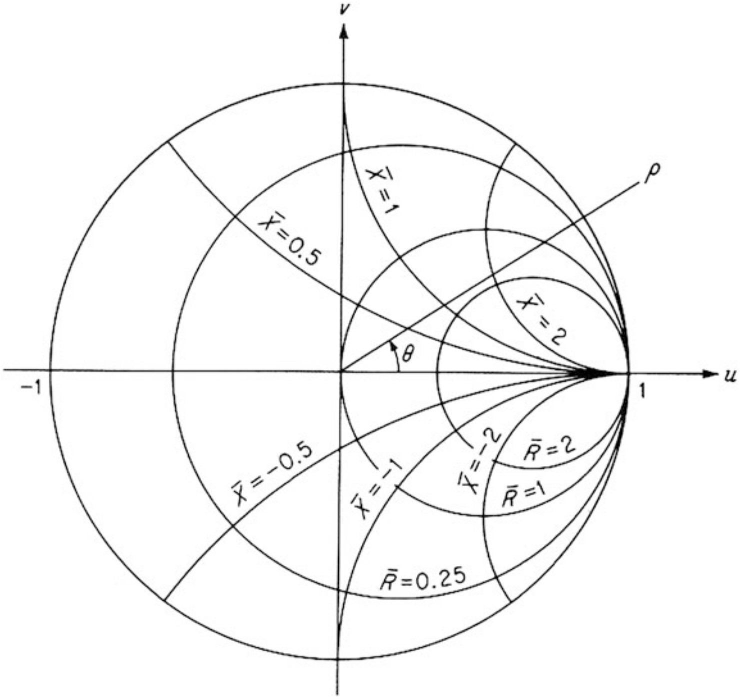


Fig. 3.3 Constant $\overline{R}$ and $\overline{X}$ circles in the reflection-coefficient plane. (From Collin [99, p. 205]. Reprinted by permission of McGraw-Hill Book Co.) The bars over letters in this figure do not mean complex conjugate. Collin calls the normalized load impedance $\overline{Z}$ instead of z and writes $\overline{Z} = \overline{R} + i\overline{X}$

Next let us consider an example of the use of the Smith chart from Collin [99, pp. 212–215]. The general idea is to see what various network elements do to a given point Γ. The goal is to finish with $\Gamma = 0$ so that load is matched. First we need a few more definitions. Let the normalized input admittance

$$y_{\text{in}} = z_{\text{in}}^{-1} = (1-\Gamma)(1+\Gamma)^{-1} = G + iB,$$

where $G =$ **conductance** and $B =$ **susceptance**.

A **stub** is a short length of transmission line with a short circuit at its end. Consider the double-stub tuner pictured in Fig. 3.5. Let P_1 on the Smith chart in Fig. 3.6 correspond to $y_L = G_L + iB_L =$ the normalized load admittance. The first stub at the plane *aa* in Fig. 3.5 adds a susceptance iB_1 which moves P_1 along the $G =$ constant circle to P_2 in Fig. 3.6. Then we get to the second stub at the plane *bb* in Fig. 3.5 by moving in the Γ-plane from P_2 to P_3 in Fig. 3.6, through the angle $2\beta d = 4\pi d/\lambda$, clockwise. The point P_4 is obtained by moving along $G =$ constant. In order to make $P_4 = 0$ and achieve matching, the point P_3 must lie on the circle $G = 1$. You can figure out graphically the value B_1 must have to do this. In fact,

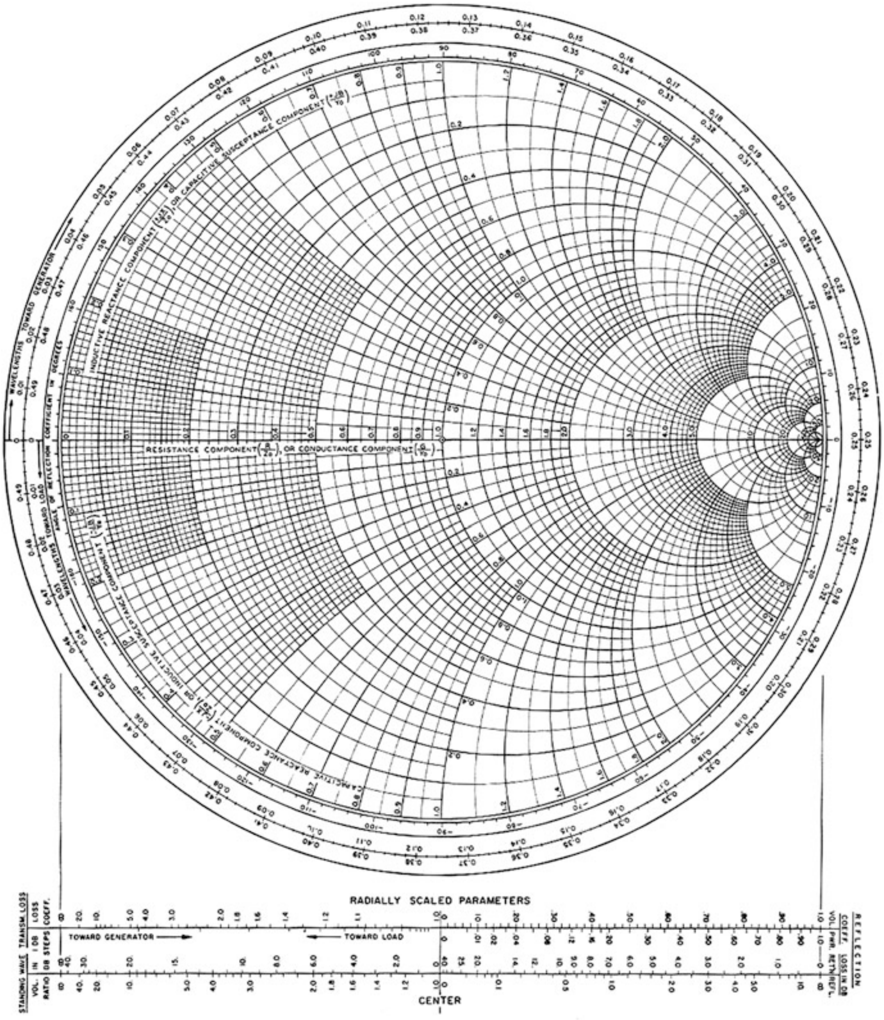
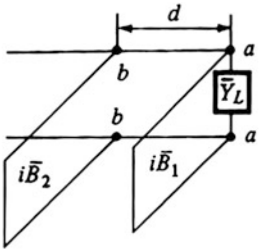


Fig. 3.4 The Smith Chart. (Reproduced by permission of Phillip H. Smith, Analog Instruments Co.)

Fig. 3.5 Double-stub tuner. (From Collin [99, p. 212]. Reproduced by permission of McGraw-Hill Book Co.). Collin puts bars over variables to denote that they correspond to normalized input impedance and not to denote a complex conjugate



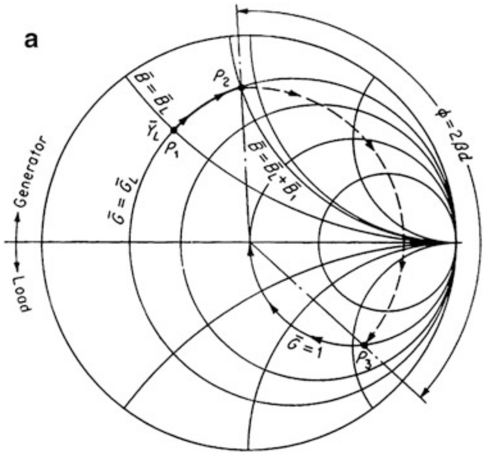


Fig. 3.6 (**a**) Representation of the operation of a double-stub tuner. Figure continued in Fig. 3.7 (From Collin [99, pp. 213]. Reproduced by permission of McGraw-Hill)

there may be two solutions or no solution, depending on the value of G_L and d. See Figs. 3.6 and 3.7 from Collin [99, pp. 213–214]. It turns out that one can always obtain matching with three stubs at fixed positions (see Collin [99, pp. 214–217]).

Note that z or y lie in the right half-plane, rather than in the Poincaré upper half-plane H. To put y in H multiply it by i:

$$y = G + iB \mapsto iy = -B + iG.$$

Then the first stub in the double-stub tuner transforms iy to $iy + B_1$. This means that the stub corresponds to the following matrix in $SL(2,\mathbb{R})$:

$$\begin{pmatrix} 1 & B_1 \\ 0 & 1 \end{pmatrix}.$$

In general, two-port circuits correspond to 2×2 matrices. The terminology does not appear to be totally standardized. But we shall describe the correspondence given in Helton [286]. The two-port circuit has the diagram shown in Fig. 3.8. For a linear two-port circuit, the **impedance matrix** $Z \in \mathbb{C}^{2\times 2}$ satisfies $ZI = V$, where

$$I = \begin{pmatrix} I_1 \\ I_2 \end{pmatrix} \text{ and } V = \begin{pmatrix} V_1 \\ V_2 \end{pmatrix}.$$

In this section we shall use $\mathbb{I}$ rather than I as the 2×2 identity matrix. The **scattering matrix** for the circuit is $S = (\mathbb{I} - Z)(\mathbb{I} + Z)^{-1}$. And the **chain matrix** ч (the Russian letter cha or che) is defined by

$$\text{ч}\begin{pmatrix} V_2 \\ -I_2 \end{pmatrix} = \begin{pmatrix} V_1 \\ I_1 \end{pmatrix}.$$

The **power consumption** is $P = {}^t\overline{V} I =$ the hermitian product of the two vectors. The **admittance matrix** is $Y = Z^{-1}$. In Fig. 3.9 we list the chain matrices for some common two-ports.

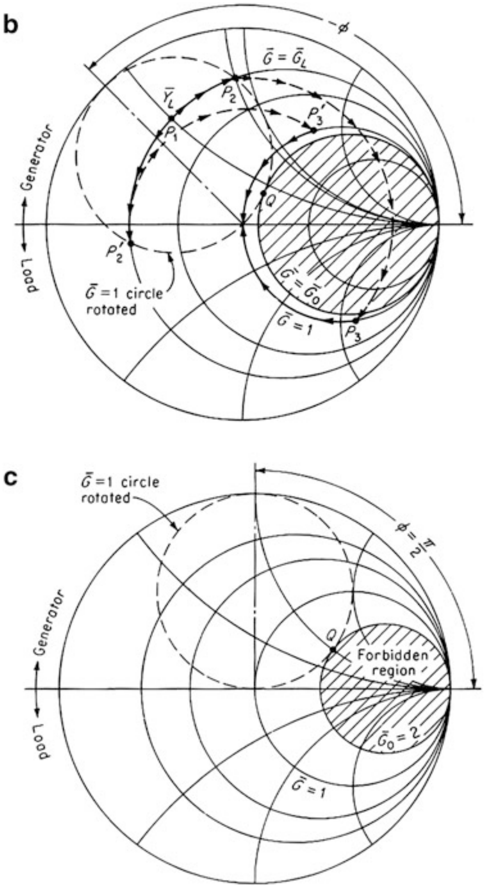


Fig. 3.7 (**b**) Determination of required susceptance for the first stub in a double-stub tuner; (**c**) Range of load impedance that cannot be matched when $d = \lambda/8$. (From Collin [99, pp. 213–214]. Reproduced by permission of McGraw-Hill)

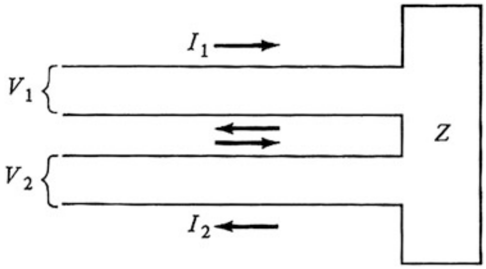


Fig. 3.8 A two-port circuit

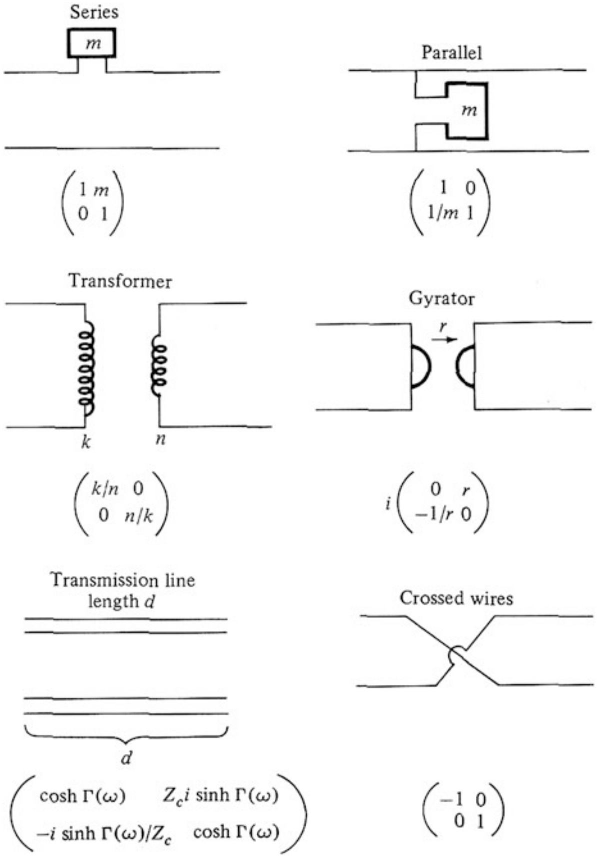

Series
m
$\begin{pmatrix} 1 & m \\ 0 & 1 \end{pmatrix}$
Parallel
m
$\begin{pmatrix} 1 & 0 \\ 1/m & 1 \end{pmatrix}$
Transformer
k
n
$\begin{pmatrix} k/n & 0 \\ 0 & n/k \end{pmatrix}$
Gyrator
r
$i\begin{pmatrix} 0 & r \\ -1/r & 0 \end{pmatrix}$
Transmission line length d
d
$\begin{pmatrix} \cosh\Gamma(\omega) & Z_c i \sinh\Gamma(\omega) \\ -i\sinh\Gamma(\omega)/Z_c & \cosh\Gamma(\omega) \end{pmatrix}$
Crossed wires
$\begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$

Fig. 3.9 Chain matrices for some common two-ports. (Based on an illustration in Helton [286], by permission of the American Mathematical Society)

3.1.3 $SL(2,\mathbb{R})$ as a Lorentz-Type Group

There is yet one more realization of the group $SL(2,\mathbb{R})$ that we should mention. This connects the group with Lorentz-type groups that are of great interest in physical applications. The connection is developed in the following exercise.

Exercise 3.1.13. Let

$$W = \left\{ X(x_0,x_1,x_2) = \begin{pmatrix} x_0+x_1 & x_2 \\ x_2 & x_0-x_1 \end{pmatrix} \middle| \; x_i \in \mathbb{R}, 0 \le i \le 2 \right\},$$

and $e_0 = X(1,0,0)$, $e_1 = X(0,1,0)$, $e_2 = X(0,0,1)$. For each g in $SL(2,\mathbb{R})$, let $s(g)$ denote the matrix of the linear transformation of W given by sending x in W to

$g\,x\,{}^t g$, using the basis $\{e_0, e_1, e_2\}$. Show that $s : SL(2,\mathbb{R}) \to G_+(2)$ is a continuous homomorphism onto $G_+(2) =$ the connected component of the identity in the three-dimensional Lorentz group $O(1,2)$. Prove that the kernel of s is $\{\pm I\}$. Here by kernel, we mean the inverse image of the identity. The Lorentz group $O(1,2)$ is defined to consist of all real 3×3 matrices which preserve the quadratic form $Qx = x_0^2 - x_1^2 - x_2^2$.

Hint. *Note that* $\det(X(x_0, x_1, x_2)) = x_0^2 - x_1^2 - x_2^2$ *is invariant under* $s(g)$. *You can find more details in Sugiura [648, pp. 206–207].*

Alternative Discussion of the Preceding Exercise. Consider a 2×2 matrix A. *If* e_1 and e_2 are basis vectors and $Ae_i = a_{1i}e_1 + a_{2i}e_2$, then we obtain the second symmetric power, denoted $Sym^2(A)$, by considering symmetric products (analogous to the alternating products appearing in the theory of differential forms):

$$\begin{aligned}
Ae_1 \cdot Ae_1 &= a_{11}^2 e_1^2 + 2a_{11}a_{21}e_1e_2 + a_{21}^2 e_2^2, \\
Ae_2 \cdot Ae_2 &= a_{12}^2 e_1^2 + 2a_{12}a_{22}e_1e_2 + a_{22}^2 e_2^2, \\
Ae_1 \cdot Ae_2 &= a_{11}a_{12}e_1^2 + (a_{11}a_{22} + a_{21}a_{12})e_1e_2 + a_{21}a_{22}e_2^2.
\end{aligned}$$

So the matrix of $Sym^2(A)$, using the basis e_1^2, e_1e_2, e_2^2, is:

$$\begin{pmatrix} a_{11}^2 & a_{11}a_{12} & a_{12}^2 \\ 2a_{11}a_{21} & a_{11}a_{22} + a_{21}a_{12} & 2a_{12}a_{22} \\ a_{21}^2 & a_{21}a_{22} & a_{22}^2 \end{pmatrix}.$$

This is a representation; i.e., $Sym^2(gh) = Sym^2(g)Sym^2(h)$.

Next note that we can view $SL(2,\mathbb{R})$ as the group $Sp(1,\mathbb{R})$, the **symplectic group** of 2×2 real matrices preserving the alternating form J; i.e.,

$${}^t g \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} g = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} = J.$$

Now

$$Sym^2(J) = \begin{pmatrix} 0 & 0 & 1 \\ 0 & -1 & 0 \\ 1 & 0 & 0 \end{pmatrix} \text{ which is equivalent to } \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}.$$

Here equivalence means by change of basis.

So now we see that the representation $Sp(1,\mathbb{R}) = SL(2,\mathbb{R}) \xrightarrow{Sym^2} SO(2,1)$, provided that $Sym^2({}^t g) = {}^t Sym^2(g)$. To make this true, replace $Sym^2(A)$ by the equivalent matrix obtained by multiplying the second row of $Sym^2(A)$ by $1/\sqrt{2}$ and the second column by $\sqrt{2}$.

3.2 Harmonic Analysis on H

> Wie man u.a. in den Werken von Euler und Gauss feststellen kann, haben in früheren Jahrhunderten die Mathematiker nicht nur allgemeine Theoreme veröffentlicht, sondem dazu auch Beispiele, die ihnen wohl Vergnügen machten und den Leser weiter belehrten.
>
> ...
>
> Mir selber erscheint es allerdings nach den entscheidenden Ergebnissen von Gödel und Cohen geraten, mit dem ungehemmten axiomatischen Verfahren weiterhin vorsichtig zu sein. So hat mir schon immer die Anwendung des Auswahlaxioms widerstrebt, und ich fühle mich sogar jedesmal auf meinen Füssen sicherer, wenn ich auf zwei verschiedenen Wegen einsehen kann, dass zweimal zwei gleich vier ist.[2]
>
> —From Siegel [600, Vol. IV, p. 33].

3.2.1 Introduction

Complete sets of eigenfunctions of the non-Euclidean Laplacian

$$\Delta = y^2\left(\partial^2/\partial x^2 + \partial^2/\partial y^2\right)$$

can be found by separation of variables in either rectangular or geodesic polar coordinates. For rectangular coordinates, the spectral resolution of Δ reduces to the inversion formula for the Kontorovich–Lebedev transform which was discussed in Exercise 2.2.8 of Sect. 2.2. This result was used by Kontorovich and Lebedev [367–369] in the late 1930s to solve various boundary value problems in mathematical physics, such as Dirichlet's problem for a wedge (see Exercise 3.1.8). For geodesic polar coordinates, the spectral decomposition of Δ boils down to the inversion formula for the Mehler–Fock transform. This is due to Mehler [462] in 1881 and was proved by Fock [181] in 1943. One motivation for studying the Mehler–Fock transform was the need to solve such physical problems as the computation of the electrostatic field due to a thin charged conductor in the shape of the surface of the region bounded by two intersecting spheres (see Exercise 3.2.23).

After considering these two different reductions of the spectral decomposition of Δ to that of certain one-variable singular Sturm–Liouville operators, we shall consider another approach due to Helgason [276, 280]. This result will then be used to find the fundamental solution of the heat equation or heat kernel on H.

---

[2] As one can see from the works of Euler and Gauss, mathematicians of previous centuries not only published general theorems but also examples which they might have enjoyed and which might have educated the reader.

...

I believe that after the decisive results of Gödel and Cohen one should be cautious with unrestrained axiomatic techniques. I was always reluctant to use the axiom of choice and am always more comfortable when I have two different techniques to see that $2\times 2 = 4$.

Then we shall discuss the central limit theorem on H, which was first proved by Karpelevich et al. [337]. This can be used to study the power reflected by random inhomogeneities in a long transmission line, as was first noted by Gertsenshtein and Vasil'ev [208]. Such work is related to investigations of Clerc and Roynette [92], Faraut [173], Heyer [294], Keller and Papanicolaou [349], Papanicolaou [509], and Trimèche [685]. There are many applications of such central limit theorems; e.g., in demography, learning theory, and atomic physics (see Cohen [96], LePage [412], and Hantsch and von Waldenfels [248]).

There are also many applications of the fundamental solution of the heat equation or heat kernel. Rosenberg [547] considers the heat kernel on compact Riemannian manifolds and applies it to Atiyah–Singer index theory.

Thus, by the end of this section, we shall have discussed three versions of harmonic analysis on H, along with a few applications. In 1947, V. Bargmann essentially obtained the analogue of such Fourier decompositions for the group $SL(2,\mathbb{R})$ itself (see Bargmann [25]). We will do little more than mention this work, along with a few of the implications for quantum mechanics, at the end of the section.

References for this section include Dym and McKean [147], Gangolli [191–194], Helgason [276–284], Lang [389], Lebedev [401,402], Rühl [550,551], and Sugiura [648].

3.2.2 *Harmonic Analysis on H in Rectangular Coordinates*

Harmonic analysis on H in rectangular coordinates x, y will be discussed first. These coordinates are the natural ones to use in the study of modular forms (automorphic forms for $SL(2,\mathbb{Z})$), as we shall see in later sections of this chapter. In order to carry out harmonic analysis on H in rectangular coordinates, one must be familiar with the eigenfunctions of the non-Euclidean Laplacian that transform nicely under translation. More precisely, we are seeking functions $f(z)$ which we call **k-Bessel functions** associated to $a \in \mathbb{R}$ and having the following properties:

$$\begin{aligned}&\text{(i) } f(z+u) = \exp(2\pi i a u) f(z), \quad \text{for all} \quad u \in \mathbb{R}, z \in H,\\ &\text{(ii) } \Delta f = \lambda f, \\ &\text{(iii) } |f(z)| \le C y^p \text{ for some positive constants } C,\ p.\end{aligned} \tag{3.14}$$

We call these functions "k-Bessel functions" because they are closely related to the classical ***K*-Bessel functions** (also called Bessel functions of imaginary argument, Macdonald's functions, and modified Bessel functions of the third kind).

In order to build up such functions, we make use of the simpler function $f(z) = (\operatorname{Im} z)^s = y^s$, which we call the **power function**. It is easily seen that this power function is an eigenfunction of the non-Euclidean Laplacian; more precisely:

$$\Delta y^s = s(s-1)y^s. \tag{3.15}$$

So now we define the **k-Bessel function** at $s \in \mathbb{C},\ z \in H,\ a \in \mathbb{R}$ by

$$k(s|z,a) = \int_{u=-\infty}^{+\infty} \left(\operatorname{Im}\left(\frac{1}{u-z}\right)\right)^{s} \exp(2\pi i a u)\, du. \tag{3.16}$$

Here we assume that Re $s > 0$.

Exercise 3.2.1 (K-Bessel Functions).

(a) The K-Bessel function may be defined (see Lebedev [401] or Watson [721]) by

$$K_s(z) = \frac{1}{2}\int_0^{\infty} \exp\left[-\frac{z}{2}\left(t+\frac{1}{t}\right)\right] t^{s-1}\, dt, \tag{3.17}$$

for Re $z > 0$. Show that if Re $s > 0$,

$$\Gamma(s)\int_{-\infty}^{+\infty} (q^2+x^2)^{-s} \exp(2irx)\, dx$$

$$= \begin{cases} 2\sqrt{\pi}\left|\frac{r}{q}\right|^{s-\frac{1}{2}} K_{s-\frac{1}{2}}(2|rq|), & \text{if } r,q \neq 0 \\ |q|^{1-2s}\Gamma\left(\frac{1}{2}\right)\Gamma\left(s-\frac{1}{2}\right), & \text{if } r=0, q\neq 0. \end{cases}$$

Hint. *Use* $\Gamma(s) = \int_0^{\infty} t^{s-1}e^{-t}$*, for Re* $s > 0$*, to see that*

$$\Gamma(s)\int_{-\infty}^{+\infty} (q^2+x^2)^{-s} \exp(2irx)\, dx$$

$$= \int_{-\infty}^{+\infty} \exp(2irx) \int_0^{\infty} t^{s-1}(q^2+x^2)^{-s} e^{-t}\, dt\, dx$$

$$= \int_{-\infty}^{+\infty} \exp(2irx) \int_0^{\infty} u^{s-1} \exp\left(-u(q^2+x^2)\right)\, du\, dx.$$

Then complete the square

$$-ux^2 + 2irx - uq^2 = -u\left[(x-ir/u)^2 + q^2 + (r/u)^2\right].$$

Now change variables via $w = x - ir/u$ *and use the fact that*

$$\int_{-\infty}^{+\infty} \exp(-w^2)dw = \sqrt{\pi}.$$

(b) Deduce from (a) that

$$k(s|z,a/\pi) = 2\sqrt{\pi}\exp(2iax)\Gamma(s)^{-1}|a|^{s-\frac{1}{2}}\sqrt{y}K_{s-\frac{1}{2}}(2|ay|), \tag{3.18}$$

when $a \neq 0$, $z = x + iy \in H$, $\operatorname{Re} s > 0$. Show also that $k(s|z,0) = y^{1-s}\Gamma\left(\frac{1}{2}\right)\Gamma\left(s-\frac{1}{2}\right)/\Gamma(s)$.

(c) Show that

$$\Delta_z k(s|z,a) = s(s-1)k(s|z,a)$$

and

$$k(s|z+x,a) = \exp(2\pi iax)k(s|z,a).$$

Exercise 3.2.2 (Asymptotics and Functional Equations of K-Bessel Functions).

(a) **Functional Equation.** Show that $K_s(z) = K_{-s}(z)$, if $\operatorname{Re} z > 0$.

(b) **Asymptotics.** Show that if s with $\operatorname{Re} s > 0$ is fixed and $\operatorname{Re} z > 0$, then

$$K_s(z) \sim 2^{s-1}\Gamma(s)z^{-s}, \text{ as } z \to 0,$$

$$K_s(z) \sim \sqrt{\frac{\pi}{2z}}e^{-z}, \text{ as } z \to \infty.$$

(c) Show that if $x > 0$ is fixed, then

$$K_{it}(x) \sim \sqrt{\frac{2\pi}{t}}e^{-\pi t/2}\sin\left(\frac{\pi}{4} + t\log t - t - t\log\frac{x}{2}\right) \quad \text{as } t \to \infty.$$

Hint. *For the first formula in part (b), you can use the first integral formula in Exercise 3.2.1(a), after making the change of variables $v = tz$. For the second formula in part (b), see Lebedev [401, p. 123], or Olver [508]. For hints on part (c), see Exercise 3.2.9 of Sect. 3.7.*

Exercise 3.2.3 (The Negativity of Δ on H).

(a) If Δf and f are both in $L^2(H)$, show that $(\Delta f, \mathrm{f}) \leq 0$, using Green's theorem. Here $(f,g) = \int f(z)\overline{g(z)}y^{-2}\,dxdy$.

(b) Show that, in fact, $(\Delta f, f) \leq -\frac{1}{4}(f,f)$.

Hint. *This is done by Dym and McKean [147, pp. 277–278].*

(c) Show that if $s(s-1) \leq 0$, then $s = \frac{1}{2} + it$, $t \in \mathbb{R}$, or $s \in [0,1]$. Prove that if $s(s-1) < -\frac{1}{4}$, then $s \notin [0,1]$.

Exercise 3.2.4 (Separation of Variables in $\Delta f = \lambda f$, Using Rectangular Coordinates). Seek solutions to $\Delta f = \lambda f$ of the form $f(x,y) = v(x)w(y)$. This leads to two ordinary differential equations:

$$\frac{w''}{w}(y) - \frac{\lambda}{y^2} = k = -\frac{v''}{v}(x),$$

where k is the separation constant. Then $v(x) = \exp(2\pi iax)$ with $k = 4\pi^2a^2$. If we set $w(y) = y^{1/2}u(y)$, show that u satisfies

$$y^2u'' + yu' - \left(4\pi^2a^2y^2 + (\lambda + \frac{1}{4})\right)u = 0.$$

Then show that (assuming we seek solutions of at most polynomial growth) if $a = 0$, a solution is $w(y) = y^s$, $\lambda = s(s-1)$, and if $a \neq 0$, a solution is $u(y) = K_{s-\frac{1}{2}}(2\pi|a|y)$, $\lambda = s(s-1)$. See Lebedev [401, pp. 109–110], if you need hints.

Theorem 3.2.1 (Fourier Inversion on H in Rectangular Coordinates). *Suppose that $f \in C_c^\infty(H)$; i.e., f is infinitely differentiable with compact support in H. Then, setting $s = \frac{1}{2} + it$,*

$$f(z) = \frac{1}{\pi^2}\int_{a\in\mathbb{R}}\int_{\mathrm{Re}\ s=\frac{1}{2}} \hat{f}(a,s)e_{a,s}(z)\ t\sinh(\pi t)\ dt\ da$$

where

$$e_{a,s} = \begin{cases} \exp(2\pi iax)\ \sqrt{y}K_{s-\frac{1}{2}}(2\pi|a|y), & \text{if } a \neq 0, \\ y^s, & \text{if } a = 0, \end{cases}$$

and

$$\hat{f}(a,s) = \int_H f(z)\overline{e_{a,s}(z)}\ y^{-2}\ dx\ dy.$$

Proof. By the Fourier inversion formula of Sect. 1.2, this result is easily reduced to the **inversion formula of Kontorovich and Lebedev** [367–369] for functions $h(y)$, $y > 0$:

$$\begin{aligned} h(y) = & \frac{2}{\pi^2}\int_0^\infty t\sinh(\pi t)\frac{K_{it}(y)}{\sqrt{y}}\int_0^\infty h(u)\frac{K_{it}(u)}{\sqrt{u}}\ du\,dt \\ f(t) = & \frac{2t\sinh(\pi t)}{\pi^2}\int_0^\infty \frac{K_{it}(y)}{y}\int_0^\infty f(u)K_{iu}(y)\ du\,dy. \end{aligned} \tag{3.19}$$

This is true because it suffices to prove the theorem when $f(z) = \upsilon(x)w(y)$, if $z = x + iy$.

There are many ways to find the spectral measure $2\pi^{-2}t\sinh(\pi t)$ in Eq. (3.19). Three ways are given in Exercise 2.2.8 of Sect. 2.2, Exercises 3.2.16 and 3.2.21 of this section. What we call the **asymptotics/functional equations principle** gives a simpler method. We shall not prove this principle rigorously here, but instead give a plausible argument for its truth. The method comes from work of Harish-Chandra and Helgason. There is a rigorous proof using Paley–Wiener theory in Helgason [280], for the special case under consideration. The generalization to symmetric spaces of semisimple Lie groups has been obtained. But no one seems to have developed the principle as a general method in spectral theory. See Exercise 3.2.6, the discussion of the Mehler–Fock inversion formula (3.26), (3.27), and Volume II [667] for more examples of the asymptotics/functional equations principle. From Exercise 3.2.2 above, we believe that the asymptotic behavior of K_{it} as y approaches 0 is given by

$$K_{it}(y) \sim 2^{it-1}\Gamma(it)y^{-it} + 2^{-it-1}\Gamma(-it)y^{it} \quad as\; y \to 0^+. \tag{3.20}$$

This can also be proved using the relation of K to the I-Bessel function and the power series for I. Note that the line of integration in the Fourier inversion integral is the line fixed by the functional equation of $K_s = K_{-s}$. And the two terms in Eq. (3.20) come from the functional equation.

To prove Eq. (3.19), one wants to see that the kernel

$$W_R(x,y) = \frac{1}{\pi^2}\int_{-R}^{+R} t\ \sinh(\pi t)\,\frac{K_{it}(x)K_{it}(y)}{\sqrt{xy}}\,dt$$

approaches $\delta(x-y)$ as $R \to \infty$. Since our problem is invariant under $SL(2,\mathbb{R})$, it is reasonable to expect that it suffices to consider only $x,y \sim 0$. But then formula (3.20) and

$$\Gamma(it)\Gamma(-it) = \pi(t\sinh(\pi t))^{-1} \tag{3.21}$$

(see Lebedev [401, p. 3]) imply that

$$W_R(x,y) \sim \frac{1}{2\pi}\int_{-R}^{+R} y^{-\frac{1}{2}-it}x^{-\frac{1}{2}+it}\,dt, \quad x,y \to 0^+. \tag{3.22}$$

And the right-hand side of Eq. (3.22) is a Dirac delta family by the Mellin inversion formula (see Exercise 1.4.1 of Sect. 1.4). So the spectral measure is chosen just to cancel out the gamma factors in Eq. (3.20).

There are also discussions of the Kontorovich–Lebedev transform in Lebedev [402] and Sneddon [609, Chap. 6]. Note that one must show that the other terms in the asymptotic expansion (3.20) do not contribute to the inversion formula. One might also ask how one knows that there is no discrete spectrum. □

Exercise 3.2.5. Show that it does not matter which of the two inversion formulas in Eq. (3.19) you prove.

Exercise 3.2.6. Can the asymptotics/functional equations principle be used to compute the spectral measure for the Hankel integral formula of Exercise 2.2.7(b) in Sect. 2.2? You want to show that:

$$\int_0^R rJ_\nu(rx)J_\nu(ry)\;r\,dr \sim \delta(x-y) \quad \text{as } R \to \infty,$$

using

$$J_\nu(x) \sim \sqrt{\frac{2}{\pi x}} \cos\left(x - \frac{\nu\pi}{2} - \frac{\pi}{4}\right), \quad \text{as } x \to \infty,\ \nu > -\frac{1}{2},$$

and

$$J_\nu(-x) = \exp(\nu\pi i) J_\nu(x).$$

Exercise 3.2.7 (Some Kontorovich–Lebedev Transforms).

(a) Show that the integral formula in Exercise 3.2.1(a) implies

$$K_{it}(z) = \int_0^\infty \exp[-z\cosh u] \cos(tu)\, du, \quad \operatorname{Re} z > 0,\ t \text{ arbitrary}.$$

Use this to deduce that

$$\frac{\pi}{2} \exp(-z\cosh u) = \int_0^\infty K_{it}(z) \cos(tu)\, dt, \quad \operatorname{Re} z > 0.$$

(b) Show that when $\operatorname{Re}(r) > |\operatorname{Re}(s)|$,

$$\int_0^\infty y^{r-1} K_s(y)\, dy = 2^{r-2} \Gamma\left(\frac{r+s}{2}\right) \Gamma\left(\frac{r-s}{2}\right).$$

Again you can use the integral formula in Exercise 3.2.1(a). For more hints, see Exercise 3.6.4 of Sect. 3.6 (the section in which we apply this result).

Note. You can find more examples in Erdélyi [165, Vol. II, pp. 125–153, 175–177], and Oberhettinger [499].

Exercise 3.2.8 (An Application of the Kontorovich–Lebedev Transform). Find the harmonic function in the wedge in Fig. 3.10 taking on given values on the boundary of the wedge; i.e., solve

$$\begin{cases} \Delta u(r,\phi,z) = 0 \quad \text{in } 0 < r < \infty,\ |\phi| \le a,\ z \in \mathbb{R}, \\ u(r,-a,z) = f(r,z), \\ u(r,a,z) = g(r,z). \end{cases}$$

Here (r,ϕ,z) denote cylindrical coordinates in $\mathbb{R}^3$ and f,g are given functions. Obtain solutions of

$$\Delta u = \frac{\partial^2 u}{\partial r^2} + \frac{1}{r}\frac{\partial}{\partial r} + \frac{1}{r^2}\frac{\partial^2 u}{\partial \phi^2} + \frac{\partial^2 u}{\partial z^2} = 0$$

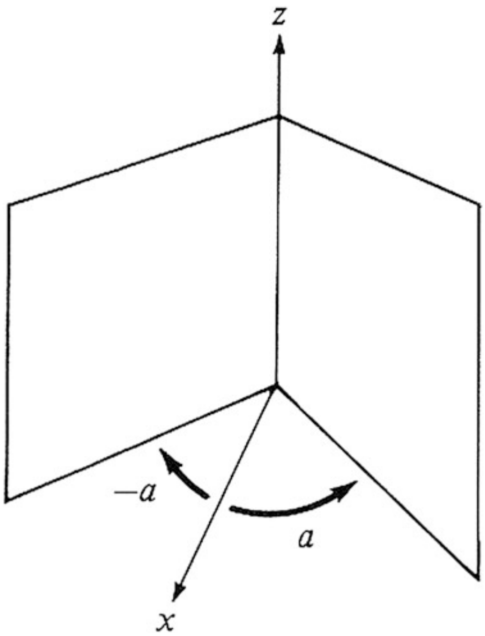


Fig. 3.10 A wedge

of the form $u(r,\phi,z) = \exp(\pm\phi t \pm icz)R(r)$. Then $R(r)$ satisfies

$$R'' + r^{-1}R' + (t^2r^{-2} - c^2)R = 0.$$

One solution is $K_{it}(r|c|)$. Use the Kontorovich–Lebedev transform to construct u from this. You can find the answer in Sneddon [609, pp. 363–366] or Lebedev [401, pp. 150–153].

3.2.3 Harmonic Analysis on H in Geodesic Polar Coordinates

Next we consider harmonic analysis on H in **geodesic polar coordinates** (r,u) given by

$$z = x + iy = k_u e^{-r} i, \quad \text{with} \quad k_u = \begin{pmatrix} \cos u & \sin u \\ -\sin u & \cos u \end{pmatrix}. \tag{3.23}$$

In Sect. 3.1, after formula (3.8), we noted that as u runs from 0 to 2π, and r runs from 0 to ∞, the upper half-plane is covered once. Recall also that we computed ds^2, $d\mu$, and Δ in geodesic polar coordinates in Exercise 3.1.11 of Sect. 3.1.

In order to do harmonic analysis on H in geodesic polar coordinates, one must have some knowledge of the associated Legendre functions $P^a_{-\frac{1}{2}+it}$. These are essentially the eigenfunctions f of Δ on H which transform under k_u defined in Eq. (3.23) according to the law

$$f(k_u z) = \exp(2iau)f(z), \quad z \in H,\ u \in \mathbb{R}. \tag{3.24}$$

Using the power function y^s to construct functions satisfying Eq. (3.24), as we did in formula (3.16), we define the **associated spherical function**

$$h(s|z,a) = \frac{1}{2\pi}\int_0^{2\pi} \operatorname{Im}(k_{-t}z)^s \exp(2iat)\,dt. \tag{3.25}$$

This turns out to be an associated Legendre function as the following exercise shows. If $a = 0$ in Eq. (3.25), we are looking at a **zonal spherical function**, invariant under rotation, which is a Legendre function.

Exercise 3.2.9 (Associated Legendre Functions). The associated Legendre function $P_s^a(z)$ can be defined for $\operatorname{Re} z > 0$ (see Lebedev [401, p. 199]) by

$$P_s^a(z) = \frac{\Gamma(s+a+1)}{2\pi\Gamma(s+1)}\int_0^{2\pi}\left[z+\sqrt{z^2-1}\cos u\right]^s \exp(iau)\,du.$$

Show that

$$h(s|z,a) = \exp(2iau)\frac{\Gamma(1-s)}{\Gamma(1-s+2a)}P_{-s}^{2a}(\cosh r), \quad \text{if} \quad z = k_u e^{-r} i.$$

Hint. *Note that* $\operatorname{Im}(k_u e^{-r} i) = [\cosh r + \cos(2u)\sinh r]^{-1}$.

Exercise 3.2.10 (Asymptotics and Functional Equation). Define the **Legendre function** by $P_s = P_s^0$.

(a) **Functional Equation.**
Show that if $\operatorname{Re} z > 0$, then $P_{-s}(z) = P_{s-1}(z)$.

(b) **Asymptotics.**
Show that if $\operatorname{Re} z > 0$, $\operatorname{Re} s > -\frac{1}{2}$, then

$$P_s(z) \sim \frac{1}{\sqrt{\pi}}\frac{\Gamma(s+\frac{1}{2})}{\Gamma(s+1)}(2z)^s, \quad \text{as } z \to \infty.$$

Hint. *Use the integral formula in Exercise 3.2.9 and make the change of variables* $x = \tan(u/2)$. *That leads to the formula*

$$P_s(z) = \frac{1}{\pi}\int_{-\infty}^{\infty}\left\{z+\sqrt{z^2-1}\left(\frac{2}{x^2+1}-1\right)\right\}^s \frac{1}{x^2+1}dx.$$

This integral approaches

$$\frac{(2z)^s}{\pi}\int_{x\in\mathbb{R}}(x^2+1)^{-s-1}dx=\frac{(2z)^s}{\pi}B\left(\frac{1}{2},s+\frac{1}{2}\right),\quad \text{as}\quad z\to\infty,$$

*where $B(a,b) =$ the **beta function** $=\Gamma(a)\Gamma(b)/\Gamma(a+b)$ (see Lebedev [401, Chap. 1]).*

Exercise 3.2.11 (Separation of Variables in $\Delta f=\lambda f$ in Geodesic Polar Coordinates). Write $f(k_u e^{-r}i)=v(r)w(u)$ with $(u,r)=$ geodesic polar coordinates in H and solve $\Delta f=\lambda f$. You will obtain

$$\frac{\sinh r}{v(r)}\left[\sinh r\, v'(r)\right]'-\lambda(\sinh r)^2=k=-\frac{w''(u)}{w(u)},$$

where k is the separation constant. Set $w(u)=\exp(iau)$, $a\in\mathbb{Z}$. Then $k=a^2$ and v must be a solution of

$$\sinh r\ \left(\sinh r\, v'(r)\right)'-\left(\lambda\sinh^2 r+a^2\right)v(r)=0.$$

Set $x=\cosh r$, $V(x)=v(r)$. Show that V must satisfy the ODE

$$(1-x^2)V''-2xV'+\left(\lambda+\frac{a^2}{x^2-1}\right)V=0.$$

Set $\lambda=s(s-1)$ and obtain a solution $V(x)=P^a_{-s}(x)$ (see Lebedev [401, p. 214]).

The functions $P^a_{-\frac{1}{2}+it}(x)$ were first considered by Mehler [462]. They came to be called conical functions because they arise in physical problems involving cones. We call them spherical functions because they are a special case of the (associated) spherical functions attached to any symmetric space, examples of which will be considered in Volume II [667]. The Legendre functions can also be viewed as Gauss hypergeometric functions ${}_2F_1$ (see Lebedev [401, p. 165]). This gives the power series in the next exercise by definition. We will be most interested in the zonal spherical functions (the case $a=0$).

Exercise 3.2.12 (A Few More Properties of Associated Legendre Functions).

(a) Show that if $|z-1|<2$ we have

$$\begin{aligned}P_s^0(z)=P_s(z)&=\sum_{k=0}^{\infty}\frac{(-s)_k(s+1)_k}{(k!)^2}\left(\frac{1-z}{2}\right)^k\\&={}_2F_1\left(-s,s+1;1;\frac{1-z}{2}\right),\end{aligned}$$

where

$$(s)_k = \begin{cases} 1, & \text{if } k = 0; \\ s(s+1)\cdots(s+k-1), & \text{if } k = 1, 2, \ldots. \end{cases}$$

and ${}_2F_1$ = the **Gauss hypergeometric function** (see Lebedev [401]).

(b) Show that if $a = 0, 1, 2, \ldots$, then $P_s^a(z) = (z^2-1)^{a/2} D^a P_s(z)$, if $D = d/dz$.

(c) Show that $P_s^a(z) = P_s^{-a}(z)\Gamma(s+a+1)/\Gamma(s-a+1)$ for $a = 0, 1, 2, \ldots$.

(d) Show that if

$$I = \int_1^z$$

then

$$P_s^{-a}(a) = (z^2-1)^{-a/2} I^a P_s(z).$$

Note that $P_\nu^a(x)$ is a polynomial when both a and ν are nonnegative integers; in fact, they are the functions that arose in the solution of the analogue of Exercise 3.2.11 for the sphere in Chap. 2. We will not have anything to say about Legendre functions of the second kind until Sect. 3.7 (but see Hermann [292, Chap. 3]). Another reference for conical and spherical functions is Robin [543].

Exercise 3.2.13 (Laplace Transforms Relating *K* and *P*). Prove that the following integral formulas hold.

(a)

$$\int_0^\infty \exp(-pt) t^{u-1/2} K_{\nu+1/2}(t)\, dt = \sqrt{\frac{\pi}{2}}\, \Gamma(u-\nu)\Gamma(u+\nu+1)(p^2-1)^{-u/2} P_\nu^{-u}(p),$$
$$\text{if} \quad \mathrm{Re}(u+\nu) > -1, \;\; \mathrm{Re}(u-\nu) > 0, \;\; \mathrm{Re}\, p > -1.$$

(b)

$$\int_1^\infty P_{\nu-\frac{1}{2}}(u) \exp(-ut)\, du = \sqrt{\frac{2}{\pi t}}\, K_\nu(t), \quad \text{if} \quad \mathrm{Re}\, t > 0.$$

Theorem 3.2.2 (Harmonic Analysis on H in Geodesic Polar Coordinates). *Let (r, u) denote geodesic polar coordinates as in formula (3.23). And define*

$$\varepsilon_{a,t}(k_u e^{-r} i) = \exp(iau) P^a_{-\frac{1}{2}+it}(\cosh r).$$

Then, if $f \in C_c^\infty(H)$ = the infinitely differentiable functions with compact support,

$$f(z) = \frac{1}{2\pi} \sum_{a\in\mathbb{Z}} (-1)^a \int_{t\in\mathbb{R}^+} \tilde{f}(a,t)\varepsilon_{-a,t}(z)\, t\tanh(\pi t)\, dt.$$

where

$$\tilde{f}(a,t) = \int_H f(z)\varepsilon_{a,t}(z)\, d\mu.$$

Proof. We show in Exercise 3.2.14 that it suffices to prove the **inversion formula for the Mehler–Fock transform** (see Mehler [462] and Fock [181]):

$$g(v) = \int_0^\infty t\tanh(\pi t)\; P_{-\frac{1}{2}+it}(v) \int_1^\infty g(w)P_{-\frac{1}{2}+it}(w)\, dw\, dt, \tag{3.26}$$

$$f(w) = w\tanh(\pi w) \int_1^\infty P_{-\frac{1}{2}+iw}(u) \int_0^\infty P_{-\frac{1}{2}+it}(w) f(t)\, dt\, du. \tag{3.27}$$

There are discussions of this result in Exercises 3.2.15 and 3.2.17 below. We can also obtain Eqs. (3.26) and (3.27) from the asymptotics/functional equations principle which we used to discuss Theorem 3.2.1 and the Kontorovich–Lebedev inversion formula. For by Exercise 3.2.10, if $|\arg x| < \pi$, then

$$P_{-\frac{1}{2}+it}(x) \sim \frac{\Gamma(it)}{\sqrt{\pi}\,\Gamma\left(\frac{1}{2}+it\right)}(2x)^{-\frac{1}{2}+it} + \frac{\Gamma(-it)}{\sqrt{\pi}\,\Gamma\left(\frac{1}{2}-it\right)}(2x)^{-\frac{1}{2}-it} \quad \text{as } x\to\infty$$

for fixed real t. This can also be proved by noting that the Legendre function is a Gauss hypergeometric function (see Lebedev [401]). Helgason [280, pp. 62–82], gives a discussion of the complete asymptotic expansion of the spherical function and its application to Fourier inversion on the symmetric space via Paley–Wiener theory.

Our goal is to show that the kernel

$$V_R(x,y) = \int_0^R t\tanh(\pi t)\; P_{-\frac{1}{2}+it}(x)P_{-\frac{1}{2}+it}(y)dt$$

approaches $\delta(x-y)$, as $R\to\infty$. As for the kernel (3.22), one can argue that the $SL(2,\mathbb{R})$-invariance of the problem means that it suffices to consider x, $y\sim\infty$. Now by a standard property of the gamma function (see Lebedev [401, p. 3]) we have

$$\frac{\Gamma(it)\Gamma(-it)}{\pi\Gamma\left(\frac{1}{2}+it\right)\Gamma\left(\frac{1}{2}-it\right)} = \frac{1}{\pi t\;\tanh(\pi t)}. \tag{3.28}$$

It follows that

$$V_R(x,y) \sim \frac{1}{\pi}\int_0^R x^{-\frac{1}{2}+it} y^{-\frac{1}{2}-it} dt \quad \text{for } x,\, y\sim\infty.$$

The right-hand side of the asymptotic relation is a Dirac delta family by Mellin inversion (see Exercise 1.4.1 of Sect. 1.4). Thus the spectral measure was chosen so as to cancel out the gamma factors in the asymptotic formula for the Legendre function, as was the case for the Kontorovich–Lebedev transform.

The rest of the discussion of Theorem 3.2.2 is outlined in the next exercise. □

Exercise 3.2.14. Suppose that I and D are as in Exercise 3.2.12.

(a) Use Exercise 3.2.12 and integration by parts to show that if $f(k_u e^{-r} i) = g(x)h(u)$, with $x = \cosh r$, then

$$\begin{aligned}\tilde{f}(a,t) &= \int_{w=1}^{\infty}\int_{v=0}^{2\pi} h(v)e^{iav} g(w)(w^2-1)^{a/2} D^a P_{-\frac{1}{2}+it}(w)\,dv\,dw\\ &= (-1)^a \int_{w=1}^{\infty}\int_{v=0}^{2\pi} h(v)e^{iav} P_{-\frac{1}{2}+it}(w) D^a\left[g(w)(w^2-1)^{a/2}\right]\,dv\,dw.\end{aligned}$$

(b) Use Exercise 3.2.12 to show that

$$\begin{aligned}&\frac{1}{2\pi}\sum_{a\in\mathbb{Z}}(-1)^a \int_{t\in\mathbb{R}^+} \tilde{f}(a,t)\varepsilon_{-a,t}(z)t\tanh(\pi t)\,dt\\ &= \frac{1}{2\pi}\sum_{a\in\mathbb{Z}}(-1)^a e^{-iau}\int_{t\in\mathbb{R}^+}(-1)^a\int_{w=1}^{\infty}\int_{v=0}^{2\pi} h(v)e^{iav}\\ &\quad\times D^a\left[g(w)(w^2-1)^{a/2}\right]P_{-\frac{1}{2}+it}(w)dw(x^2-1)^{-a/2}I^a P_{-\frac{1}{2}+it}(x)t\tanh(\pi t)\,dt.\end{aligned}$$

(c) Finish the proof of Theorem 3.2.2 by using the Mehler–Fock inversion formula and the fact that $h(u)$ is represented by its Fourier series.

(d) Show that $\overline{\varepsilon_{a,t}(z)} = \varepsilon_{-a,-t}(z) = \varepsilon_{-a,t}(z)$.

Exercise 3.2.15. Can you find a proof of the Mehler–Fock inversion formula which is analogous to the proof of the Kontorovich–Lebedev inversion formula in Exercise 2.2.8 of Sect. 2.2?

Exercise 3.2.16 (Another Derivation of the Kontorovich–Lebedev Inversion Formula).

(a) Use the integral formula in Exercise 3.2.1 to see that $K_{it}(y)$ is the Fourier transform on $\mathbb{R}$ of $h_y(x) = \frac{1}{2}\exp(-y\cosh x)$; i.e., $K_{it}(y) = \widehat{h_y}(t/2\pi)$.

(b) Use the multiplication formula for the Fourier transform on $\mathbb{R}$ (see Theorem 1.2.1 of Sect. 1.2) to show that if $f(x)$ is extended to $\mathbb{R}$ by setting $f(x) = f(-x)$, then

$$\begin{aligned}&\frac{2}{\pi^2}x\sinh(\pi x)\int_{y=0}^{\infty}K_{ix}(y)\frac{1}{y}\int_{u=0}^{\infty}f(u)K_{iu}(y)\,du\,dy\\ &= \frac{1}{2\pi^2}x\sinh(\pi x)\int_{y=0}^{\infty}K_{ix}(y)\frac{1}{y}\int_{u=-\infty}^{+\infty}\hat{f}\left(\frac{u}{2\pi}\right)\exp(-y\cosh u)\,du\,dy.\end{aligned}$$

(c) Show that the Kontorovich–Lebedev inversion formula (3.19) follows from Fourier inversion on $\mathbb{R}$ (Sect. 1.2) and

$$\int_{y=0}^{\infty} e^{-y\cosh u} K_{ix}(y)\frac{1}{y}dy = \pi\frac{\cos(xu)}{x\sinh(\pi x)}.$$

This last integral formula follows from part (a) of Exercise 3.2.13, since

$$P_{\nu}^{\frac{1}{2}}(z) = \frac{\sqrt{2\pi}}{(z^2-1)^{\frac{1}{4}}}\left(\left(z+\sqrt{z^2-1}\right)^{\nu+\frac{1}{2}} + \left(z+\sqrt{z^2-1}\right)^{-\nu-\frac{1}{2}}\right)$$

(see Erdélyi et al. [164, Vol. I, p. 150]). You also need formula (3.21).

Exercise 3.2.17 (Another Derivation of the Mehler–Fock Inversion Formula). Use Exercise 3.2.13(a) to see that

$$\begin{aligned}&\int_{u\geq 1}\int_{t\geq 0} P_{-\frac{1}{2}+iw}(u)P_{-\frac{1}{2}+it}(u)g(t)\,dt\,du\\ &= \frac{\sqrt{2}}{\pi^{3/2}}\int_{v\geq 0}\frac{1}{\sqrt{v}}\int_{u\geq 1} e^{-uv}P_{-\frac{1}{2}+iw}(u)\int_{t\geq 0}K_{it}(v)g(t)\cosh(\pi t)\,dt\,du\,dv.\end{aligned}$$

Then use Exercise 3.2.13(b) and the Kontorovich–Lebedev inversion formula to prove the Mehler–Fock inversion formula.

Exercise 3.2.18 (Some Mehler–Fock Transforms).

(a) Show that if Re $a > \frac{1}{2}$, then

$$\int_1^{\infty} w^{-a}P_{-\frac{1}{2}+it}(w)\,dw = \frac{2^{a-2}}{\Gamma(a)\sqrt{\pi}}\Gamma\left(\frac{a+it-\frac{1}{2}}{2}\right)\Gamma\left(\frac{a-it-\frac{1}{2}}{2}\right).$$

(b) Prove that if Re $a > 0$, Re $b > 0$, then

$$\int_1^{\infty}\frac{\exp[-\sqrt{a^2+b^2+2abw}]}{\sqrt{a^2+b^2+2abw}}P_{-\frac{1}{2}+it}(w)\,dw = \frac{2}{\pi\sqrt{ab}}K_{it}(a)K_{it}(b).$$

There are many other examples of Mehler–Fock transforms in Erdélyi et al. [165, Vol. II, pp. 320–326]. Mathematica should be able to provide examples as well. See Vilenkin [704] and Wawrzyńczyk [722] for discussions of spherical functions attached to group representations. We will have more to say about this subject in Volume II [667].

3.2.4 The Helgason Transform on H

Here we give our third and last version of non-Euclidean harmonic analysis on H, that of Helgason [276, 280]. Suppose that $f \in C_c^\infty(H)$. Define the **Helgason transform** of f for $s \in \mathbb{C}$, $k \in SO(2)$, by

$$\mathcal{H}f(s,k) = \int_H f(z)\overline{\operatorname{Im}(k(z))^s}\, y^{-2}dxdy. \tag{3.29}$$

Set $B = K/M$, where $M = \{-I, I\}$. Then B is called the **boundary** of H. This is reasonable since B can be identified with the circle, which is just the one point compactification of the real line. Note that $\mathcal{H}f(s,k)$ depends only on the coset kM of k in B. We will see in Lemma 3.7.3 of Sect. 3.7 that the transform Eq. (3.29) is the same transform which number theorists call the Selberg transform, when f is K-invariant (see Kubota [375, p. 56]).

The boundary of H has the non-Euclidean analogues of properties of the boundary of the unit disc. For example, it is possible to generalize classical potential theory in the framework of symmetric spaces (see Helgason [276], Koranyi [370], and Volume II [667]). Eigenfunctions of invariant differential operators on H can be shown to be given by a Poisson integral over the boundary (even in the higher-rank case) as was shown by Helgason (see also Kashiwara et al. [338]). The precise statement of this result requires Sato's theory of analytic functionals or hyperfunctions. These are elements of the dual space $\mathcal{A}'(B)$ to the space $\mathcal{A}(B)$ of analytic functions on the boundary. The space $\mathcal{A}(B)$ has a natural topology which is described for $B =$ the circle as follows. Let U be an open annulus containing B and $\mathcal{A}(U) =$ the holomorphic functions on U topologized by uniform convergence on compact subsets. Identify $\mathcal{A}(B)$ with the union of all the $\mathcal{A}(U)$ (with the inductive limit topology). The eigenfunctions are then just these analytic functionals acting on the Poisson kernel, which is just the power function.

Theorem 3.2.3 (Helgason's Version of Harmonic Analysis on H).

(1) ***Fourier Inversion on H.***
If $f \in C_c^\infty(H)$, then

$$f(z) = \frac{1}{4\pi}\int_{t\in\mathbb{R}} \frac{1}{2\pi}\int_{u=0}^{2\pi} \mathcal{H}f\left(\frac{1}{2}+it, k_u\right) \operatorname{Im}(k_u z)^{\frac{1}{2}+it}\, t\tanh(\pi t)\; du\,dt,$$

where $k_u \in SO(2)$ is defined in formula (3.23).

(2) ***The Paley–Wiener Theorem.***
The map $f \mapsto \mathcal{H}f$ takes $C_c^\infty(H)$ one-to-one, onto the space of C^∞ functions $G(s,k)$ on $\mathbb{C}\times SO(2)$ which are holomorphic in s and have properties (a), (b) below:

(a) G is of ***uniform exponential type*** *R:*

$$\sup_{s\in\mathbb{C},k\in SO(2)} \left(e^{-R|\operatorname{Re} s|}(1+|s|)^N |G(s,k)|\right) < \infty \text{ for each } N \in \mathbb{Z}^+.$$

(b) G satisfies the ***functional equation***

$$\int_{u=0}^{2\pi} \mathrm{Im}(k_u z)^{\frac{1}{2}+it} G\left(\frac{1}{2}+it, k_u\right) du \\ = \int_{u=0}^{2\pi} \mathrm{Im}(k_u z)^{\frac{1}{2}-it} G\left(\frac{1}{2}-it, k_u\right) du.$$

(3) ***The Plancherel Formula.***
The map $f \mapsto \mathcal{H}f$ extends to an isometry mapping $L^2(H, d\mu)$ onto $L^2(\mathbb{R} \times K, \frac{1}{8\pi^2} t \tanh(\pi t)\; dt\, du)$, where $K = SO(2)$ is identified with the interval $(0, 2\pi]$ using the map sending k_u to u.

Proof. We shall only discuss part (1). See Helgason [276, 280] for an alternate derivation of part (1) of the theorem, as well as proofs of parts (2) and (3). We will follow Helgason's method for $SL(n,\mathbb{R})/SO(n)$ in Volume II [667].

First let us relate the transform $\tilde{f}(a,t)$ of Theorem 3.2.2 to $\mathcal{H}f\left(\frac{1}{2}+it, k\right)$. Set

$$c(a,t) = \Gamma\left(\frac{1}{2}+it+a\right)\left[2\pi\Gamma\left(\frac{1}{2}+it\right)\right]^{-1}.$$

Then, by the integral formula in Exercise 3.2.9, we have

$$\begin{aligned} \tilde{f}(a,t) &= \int_{u=0}^{2\pi}\int_{r=0}^{\infty} f(k_{-u}e^{-r}i)e^{iau} P^{a}_{-1/2+it}(\cosh r)\; \sinh r\, dr\, du \\ &= c(a,t)\int_{u=0}^{2\pi}\int_{r=0}^{\infty} f(k_{-u}e^{-r}i)e^{iau}\int_{\theta=0}^{2\pi} e^{ia\theta}\left[\mathrm{Im}\left(k_\theta e^{-r}i\right)\right]^{1/2-it}\; \sinh r\, d\theta\, dr\, du \\ &= c(a,t)\int_{\varphi=0}^{2\pi} e^{ia\varphi}\int_{u=0}^{2\pi}\int_{r=0}^{\infty} f(k_{-u}e^{-r}i)\mathrm{Im}(k_\varphi k_{-u}e^{-r}i)^{1/2-it}\; \sinh r\, dr\, du\, d\varphi. \end{aligned}$$

To obtain the last equality, we substituted $\varphi = u + \theta$. Thus

$$\tilde{f}(a,t) = c(a,t)\int_{\varphi=0}^{2\pi} \mathcal{H}f\left(\frac{1}{2}+it, k_\varphi\right) e^{ia\varphi} d\varphi.$$

By Theorem 3.2.2, Exercises 3.2.9 and 3.2.10, we have

$$\begin{aligned} f(k_{-u}e^{-r}i) &= \frac{1}{4\pi}\sum_{a\in\mathbb{Z}}\int_{t=-\infty}^{+\infty}(-1)^a c(a,t)\int_{\varphi=0}^{2\pi} e^{ia\varphi}\; \mathcal{H}f\left(\frac{1}{2}+it, k_\varphi\right) e^{-iau} P^{-a}_{-\frac{1}{2}+it}(\cosh r)\; t\tanh(\pi t)\; d\varphi\; dt \\ &= \frac{1}{4\pi}\sum_{a\in\mathbb{Z}}(-1)^a\int_{t=-\infty}^{+\infty} c(a,t)c(-a,-t) \\ &\qquad \times\int_{\varphi=0}^{2\pi} e^{ia\varphi}\; \mathcal{H}f\left(\frac{1}{2}+it, k_\varphi\right) e^{-iau}\int_{\theta=0}^{2\pi} e^{-ia\theta}\,\mathrm{Im}(k_\theta e^{-r}i)^{\frac{1}{2}+it} t\tanh(\pi t)\; d\theta\; d\varphi\; dt. \end{aligned}$$

It follows from $\Gamma(z)\Gamma(1-z)=\pi\csc(\pi z)$ that $c(a,t)c(-a,-t)=(-1)^a(2\pi)^{-2}$ (see Lebedev [401, p. 3]). Therefore

$$f(k_{-u}e^{-r}i) = \frac{1}{16\pi^3}\sum_{a\in\mathbb{Z}}\int_{t=-\infty}^{+\infty} t\tanh(\pi t)\int_{\varphi=0}^{2\pi}\int_{\theta=0}^{2\pi}\exp[ia(\varphi-u-\theta)] \times \left(\operatorname{Im}(k_\theta e^{-r}i)\right)^{\frac{1}{2}+it}\ \mathcal{H}f\left(\frac{1}{2}+it,k_\varphi\right)\,d\theta\,d\varphi\,dt.$$

Now $(\operatorname{Im}(k_\theta e^{-r}i))^{\frac{1}{2}+it}$, as a function of θ, is represented by its Fourier series, which implies

$$f(k_{-u}e^{-r}i) = \frac{1}{8\pi^2}\int_{\varphi=0}^{2\pi}\int_{t=-\infty}^{+\infty} t\tanh(\pi t)\ \left(\operatorname{Im}(k_\varphi k_{-u}e^{-r}i)\right)^{\frac{1}{2}+it}\mathcal{H}f\left(\frac{1}{2}+it,k_\varphi\right)\,dt\,d\varphi.$$

This completes the proof of part (1) of Theorem 3.2.3. □

Helgason [276, pp. 9–10], gives a much more elegant reduction of the proof of Theorem 3.2.3, part (1), to the case of $K=SO(2)$-invariant functions $f\in C_c^\infty(H)$. The idea is to imitate the discussion of Fourier inversion on $\mathbb{R}^m$ which was given in Sect. 1.2. Define the **inverse transform** $\mathcal{S}$ for functions $h:\mathbb{R}\times B\to\mathbb{C}$ by

$$\mathcal{S}h(z) = \frac{1}{8\pi^2}\int_{t\in\mathbb{R}}\int_{u=0}^{2\pi} h(t,k_u)\left(\operatorname{Im}(k_u z)\right)^{\frac{1}{2}+it}\ t\tanh(\pi t)\ dt\,du. \tag{3.30}$$

Then Helgason proves that

$$\int_H f\,\overline{\mathcal{S}\mathcal{H}g}\,d\mu = \int_H (\mathcal{S}\mathcal{H}f)\overline{g}\ d\mu, \tag{3.31}$$

and that $\mathcal{S}\mathcal{H}$ commutes with the action of $G=SL(2,\mathbb{R})$ on H. We shall discuss this proof in detail for $G=SL(n,\mathbb{R})$ in Volume II [667].

Next we want to consider the **Helgason transform of K= $SO(2)$-invariant functions.** Suppose $f(z)=f(kz)$, for all k in $SO(2)$ and z in H. Then the Helgason transform of f is

$$\begin{aligned}\hat{f}(s) = \mathcal{H}f(s,k) = \mathcal{H}f(s,I) &= \int_H f(z)y^{\overline{s}-2}\,dx\ dy\\ &= 2\pi\int_{r=0}^{\infty} f(e^{-r}i)P_{-\frac{1}{2}+it}(\cosh r)\sinh r\,dr,\end{aligned} \tag{3.32}$$

if $s=\frac{1}{2}+it$, and $I=$ the 2×2 identity matrix. The **inverse transform** $\mathcal{S}$ for $\hat{f}(s)$ is a transform which takes functions $F:\mathbb{R}\to\mathbb{C}$ onto functions $\mathcal{S}F$ with domain H, defined by

$$\mathcal{S}F(ke^{-r}i) = \frac{1}{4\pi}\int_{t\in\mathbb{R}} F(t)P_{-\frac{1}{2}+it}(\cosh r)\ t\tanh(\pi t)\ dt \tag{3.33}$$

for k in $SO(2)$ and $r>0$. Then if $f\in C_c^\infty(H)$ and $f(k(z))=f(z)$ for all $k\in SO(2)$ and $z\in H$, the **Fourier inversion formula** of Theorem 3.2.3 is

$$f=\mathcal{S}(\hat{f}). \tag{3.34}$$

This is nothing else but the Mehler–Fock inversion formula (3.26) and (3.27). Note that for rotation-invariant functions on H, the Helgason transform is essentially an inner product with the zonal spherical function $P_{-\frac{1}{2}+it}(\cosh r)$.

Exercise 3.2.19 (Convolution).

(a) Let $f,g:SL(2,\mathbb{R})=G\to\mathbb{C}$. Define the **convolution** $f*g$ by

$$(f*g)(a)=\int_G f(b)g(b^{-1}a)\,db,$$

where $db=$ a right- and left- invariant Haar measure on G. Suppose that either f or g is really a function on $K\backslash G/K$; i.e., it is a K-invariant function on H, where $K=SO(2)$, as usual. Show that $\mathcal{H}(f*g)=\mathcal{H}f\cdot\mathcal{H}g$. Deduce that $L^1\left(K\backslash G/K\right))$ is a commutative algebra under convolution.

(b) Show that you should not expect the convolution property to hold unless either f or g is K-invariant.

Hint. *In Volume II [667], we generalize this exercise to $SL(n,\mathbb{R})$, the special linear group of $n\times n$ matrices of determinant 1. The proof may be slightly clearer in the general context. It uses a property of the power function.*

Exercise 3.2.20 (Writing the Geodesic Radial Coordinate r in Terms of the Rectangular Coordinates x,y). Let $x+iy=k_u e^{-r}i$ as in formula (3.23) above. Note that this means that for some v, we have

$$M=\begin{pmatrix}1 & x\\ 0 & 1\end{pmatrix}\begin{pmatrix}\sqrt{y} & 0\\ 0 & 1/\sqrt{y}\end{pmatrix}=k_u\begin{pmatrix}\exp(-r/2) & 0\\ 0 & \exp(r/2)\end{pmatrix}k_v.$$

Deduce that

$$M\ {}^tM=k_u\begin{pmatrix}e^{-r} & 0\\ 0 & e^r\end{pmatrix}k_{-u}$$

and that

$$Tr\left(M\ {}^tM\right)=y+\left(1+x^2\right)/y=2\cosh r.$$

The following exercise concerns a transform that will appear again as part of the Selberg trace formula in Sect. 3.7.

We spoke of geodesics in the last section. Another kind of curve in H is of interest to us as well. This is the **horocycle**. The prime example is a horizontal line in H. Any image of such a line by an element of $SL(2,\mathbb{R})$ is also a horocycle. Figure 3.11 shows horocycles through the origin. The horocycles are all orthogonal to the geodesics.

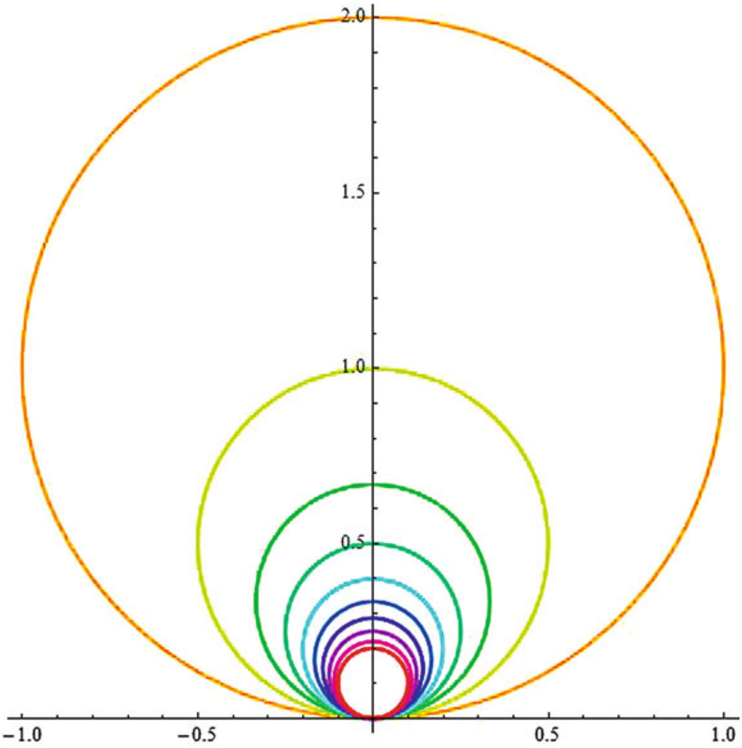


Fig. 3.11 Some horocycles in H

Exercise 3.2.21 (The Helgason Transform of K-Invariant Functions Is a Composition of Harish and Mellin Transforms).

(a) Define the **Harish transform** (also called Abel or horocycle transform) Tf of $f : H \to \mathbb{C}$ by

$$Tf(y) = \frac{1}{\sqrt{y}} \int_{x \in \mathbb{R}} f(x+iy)dx.$$

See Lang [389, pp. 69 ff]. Let Mg denote the Mellin transform of $g : \mathbb{R}^+ \to \mathbb{C}$ as in Sect. 1.4. Show that the Helgason transform of a K-invariant function f, $K = SO(2)$ (i.e., $f(kz) = f(z)$ for all $k \in K$ and $z \in H$) is given by

$$\hat{f}(s) = MTf\left(\overline{s} - \frac{1}{2}\right).$$

(b) Show that if $g : \mathbb{R}^+ \to \mathbb{C}$ and

$$G(v) = \int_{x \in \mathbb{R}} g\left(v + \frac{x^2}{2}\right)\, dx,$$

Table 3.1 Short table of Helgason transforms

$f(z) = f(k(z)) = f(e^{-r}i)$	$\hat{f}(s),\ s = \frac{1}{2}+it$
$e^{-T/4}\sqrt{2}(4\pi T)^{-3/2}\int_r^\infty \frac{be^{-b^2/4T}\,db}{\sqrt{\cosh b - \cosh r}}$	$e^{s(s-1)T}$
$(\cosh r)^{-a}, \quad a > 0$	$\sqrt{\pi}2^{a-1}\left\lvert\Gamma\left(\frac{a-\frac{1}{2}+it}{2}\right)\right\rvert^2/\Gamma(a)$
$\exp(-a\cosh r), \quad a > 0$	$2(2\pi/a)^{1/2}K_{s-\frac{1}{2}}(a)$

then

$$g(v) = \frac{-1}{2\pi}\int_{w\in\mathbb{R}} G'\left(v+\frac{w^2}{2}\right)\,dw.$$

(c) Set $f(x+iy) = g(\cosh r)$, when $x+iy$ and r are related as in Exercise 3.2.20. Show that $Tf(y) = G((y+y^{-1})/2)$, where G is the transform of g defined in part (b) and T is defined in part (a).

(d) Use (a) and (b) to deduce the inversion formula for the Helgason transform of a K-invariant function at $z = i$ (see Helgason [280], pp. 79–82).

Hint. *Mellin inversion implies that*

$$G\left(\frac{y+1/y}{2}\right) = \frac{1}{2\pi}\int_{t\in\mathbb{R}} \hat{f}(it)y^{it}\,dt.$$

Set $t = e^v$ and show that

$$G'(\cosh v) = \frac{-1}{2\pi}\int_{t\in\mathbb{R}} \hat{f}(it)\frac{t\,\sin(tv)}{\sinh(v)}\,dt.$$

Then use part (b) to see that

$$\begin{aligned} f(i) = g(1) &= \frac{-1}{2\pi}\int_{w\in\mathbb{R}} G'(1+w^2/2)\,dw \\ &= \frac{1}{4\pi^2}\int_{v\in\mathbb{R}}\int_{t\in\mathbb{R}} \hat{f}(it)\,t\,\sin(tv)\frac{\cosh(v/2)}{\sinh v}\,dt\,dv \end{aligned}$$

Here we changed variables according to $\cosh v = 1+w^2/2$, $dw = \cosh(v/2)dv$.

We give a very short table of Helgason transforms. See Table 3.1.

Exercise 3.2.22. Check Table 3.1 which lists three Helgason transforms.

Hint. *For the first line of Table 3.1 use Exercise 3.2.21 to see that if* $f(s) = e^{s(s-1)T}$, $s = \frac{1}{2}+it$, *then*

$$\begin{aligned} F(\cosh r) &= \frac{1}{2\pi}\int_{t\in\mathbb{R}} \exp\left(-\left(\frac{1}{4}+t^2\right)T\right)\exp(irt)\,dt \\ &= \frac{1}{2\sqrt{\pi T}}\exp\left(-\frac{T^2+r^2}{4T}\right). \end{aligned}$$

Also

$$\begin{aligned} f(e^{-r}i) &= \frac{-1}{2\pi}\int_{w\in\mathbb{R}} F'(\cosh r + w^2/2)\,dw \\ &= \frac{-1}{\pi}\int_r^\infty \frac{F'(\cosh v)\sinh v\,dv}{\sqrt{2(\cosh v-\cosh r)}}. \end{aligned}$$

No one seems to have evaluated this integral beyond what is in Table 3.1 (a result that would be of interest for work on the central limit theorem on H and its applications).

For the second line, see Exercise 3.2.18.

For the third line, see Exercise 3.2.13.

Exercise 3.2.23 (A Conductor That Is the Surface of Two Intersecting Spheres).

(a) Toroidal coordinates (a,b,ϕ) are defined as follows in terms of cylindrical coordinates (r,ϕ,z) (see Lebedev [401, p. 222]):

$$r = c\frac{\sinh a}{\cosh a-\cos b},\; z = c\frac{\sin b}{\cosh a - \cos b},$$

where c is a constant. Show that the surface $b=$ constant is a sphere

$$(z-c\cot b)^2 + r^2 = \left(\frac{c}{\sin b}\right)^2.$$

(b) Show that

$$\begin{aligned} \Delta f &= f_{xx}+f_{yy}+f_{zz} \\ &= \frac{a}{\partial a}\left(\frac{\sinh a}{\cosh a-\cos b}\frac{\partial f}{\partial a}\right) + \frac{\partial}{\partial b}\left(\frac{\sinh a}{\cosh a - \cos b}\frac{\partial f}{\partial b}\right) \\ &\quad + \frac{1}{(\cosh a-\cos b)\sinh a}\frac{\partial^2 f}{\partial\phi^2}. \end{aligned}$$

(c) Consider the Dirichlet problem on the domain in Fig. 3.12. Here the z-axis passes through the center of the two spheres. Choose c equal to the radius of the circle of intersection. The two spheres are given by $b=b_1$ and $b=b_2$ in toroidal coordinates. Then the Dirichlet problem is to find u with

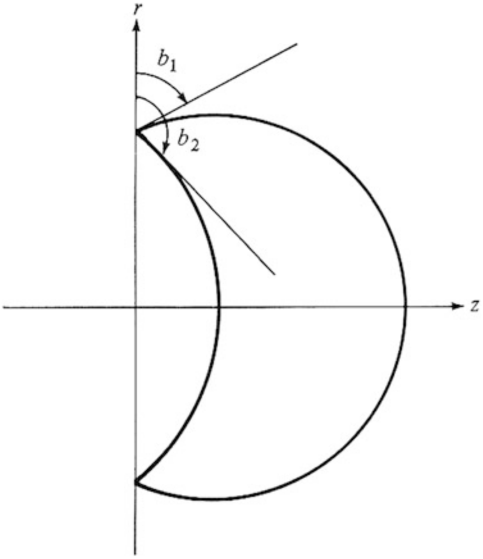


Fig. 3.12 A slice of two intersecting spheres

$$\begin{cases} \Delta u = 0, & b_1 < b < b_2, \\ u|_{b=b_j} = f_j, & j = 1,2 \end{cases}$$

Show that separation of variables leads to a solution which has the form

$$u=\sqrt{2(\cosh a - \cos b)} \int_0^\infty \frac{F_2 \sinh((b-b_1)t) + F_1 \sinh((b_2-b)t)}{\sinh((b_2-b_1)t)} P_{-\frac{1}{2}+it}(\cosh a)dt,$$

where

$$F_j(t) = t \tanh(\pi t) \int_0^\infty \frac{f_j(a)}{\sqrt{2(\cosh a - \cos b_j)}} P_{-\frac{1}{2}+it}(\cosh a) \sinh a\, da,$$

Hint. *See Lebedev [401, pp. 227–230].*

3.2.5 *The Heat Equation on* H *in Rectangular Coordinates*

The problem is to find $u = u(z,t) =$ the temperature at z in H and time t, if $u(z,0) = f(z)$ is the initial heat distribution. That is, we want to solve the heat equation on H :

$$\left. \begin{array}{c} u_t = \Delta_z u = y^2 \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \right) u(x+iy,t) \\ u(z,0) = f(z). \end{array} \right\} \tag{3.35}$$

Separation of variables leads us to consider elementary solutions of the form $u(z,t) = Z(z)T(t)$, where

$$\frac{\Delta Z}{Z} = k = \frac{T'}{T}(t).$$

Set $k = -\left(v^2 + \frac{1}{4}\right) = s(s-1)$, if $s = \frac{1}{2} + iv$. Then $T(t) = \exp\left[-\left(v^2 + \frac{1}{4}\right)t\right]$ and $Z(z) = e_{a,s}(z)$, as in Theorem 3.2.1. Then we can use the spectral decomposition of Δ in Theorem 3.2.1 to obtain

$$u(z,t) = \frac{1}{\pi^2} \int_{b\in\mathbb{R}} \int_{v\in\mathbb{R}} v \sinh(\pi v) A(b,v) e_{b,\frac{1}{2}+iv}(z) e^{-\left(v^2+\frac{1}{4}\right)t} \, dv\, db, \tag{3.36}$$

where

$$A(b,v) = \int_H f(z) \overline{e_{b,\frac{1}{2}+iv}(z)} \, y^{-2} \, dx\, dy.$$

Then Exercises 3.2.24 and 3.2.25 below show that this solution can be rewritten in the form

$$\left.\begin{array}{l} u(z,t) = f * G_t, \\ G_t(ke^{-r}i) = \frac{1}{4\pi} \int_{b\in\mathbb{R}} b \tanh(\pi b) P_{-\frac{1}{2}+ib}(\cosh r) e^{-\left(b^2+\frac{1}{4}\right)t} \, db \end{array}\right\} \tag{3.37}$$

where convolution of functions on H is induced from convolution of functions on $SL(2,\mathbb{R})$ as in Exercise 3.2.19. A function $f : H \to \mathbb{C}$ induces a function $f : SL(2,\mathbb{R}) \to \mathbb{C}$ by writing $f(g) = f(gi)$ for g in $SL(2,\mathbb{R})$, $i = \sqrt{-1}$. The function G_t in Eq. (3.37) is the **fundamental solution of the heat equation** or **heat kernel** and thus gives rise to the non-Euclidean analogue of the normal distribution discussed in Example 1.1.3 of Sect. 1.1 and Example 1.2.1 of Sect. 1.2, as well as in the discussion of the central limit theorem at the end of Sect. 1.2. We will obtain a more direct treatment of formula (3.37) using Helgason's transform, after the next two exercises. Then we will develop the non-Euclidean central limit theorem for $SO(2)$-invariant random variables.

Exercise 3.2.24.

(a) Prove that if $z = x + iy$ and $w = u + iv$ in H, then we can write Eq. (3.36) as

$$u(z,t) = \int_{w\in H} f(w) G_t(z,w) v^{-2} \, du\, dv,$$

where

$$\begin{aligned} G_t(z,w) = {} & \frac{\sqrt{vy}}{\pi^2} \int_{r\in\mathbb{R}} \exp\left(-\left(r^2 + \frac{1}{4}\right)t\right) r \sinh(\pi r) \\ & \times \int_{b\in\mathbb{R}} \exp[2\pi i b(x-u)] K_{ir}(2\pi|b|v) K_{ir}(2\pi|b|y) \, db\, dr. \end{aligned}$$

(b) Show that

$$\int_{b=0}^{\infty} \cos(xc)K_s(ax)K_s(bx)dx = \frac{\pi^2 \sec(\pi s)}{4\sqrt{ab}} P_{s-\frac{1}{2}}\left(\frac{a^2+b^2+c^2}{2ab}\right).$$

(c) Use (a) and (b) to prove

$$G_t(z,w) = \frac{1}{4\pi}\int_{b\in\mathbb{R}} e^{-(b^2+\frac{1}{4})t}\, b\tanh(\pi b)\; P_{ib-\frac{1}{2}}\left(\cosh r(z,w)\right)\, db,$$

where $r(z,w)$ is the geodesic radial coordinate [see formula (3.23)] of the point $M_w^{-1}M_z i$, if $M_z i = z$ and $M_w i = w$ for M_z, M_w in $SL(2,\mathbb{R})$. In particular, we have

$$\begin{aligned} M_z &= \begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix}\begin{pmatrix} \sqrt{y} & 0 \\ 0 & 1/\sqrt{y} \end{pmatrix} \\ &= k\begin{pmatrix} e^{-r/2} & 0 \\ 0 & e^{r/2} \end{pmatrix} k', \quad k,k' \in K, \end{aligned}$$

$y + \left(1+x^2\right)/y = 2\cosh r$, as in Exercise 3.2.20. Then use Exercise 3.2.25 to see that $2\cosh r(z,w) = (y^2+v^2+(x-u)^2)/(vy)$, if z and w are as in part (a).

(d) Prove formula (3.37).

Exercise 3.2.25.

(a) Suppose that $z = x+iy$ and use the same matrix M_z as in part (c) of Exercise 3.2.24. Similarly, define M_w for $w = u+iv$. Show that

$$M_z M_w^{-1} i = \left(\frac{xv-uy}{v}\right) + \frac{y}{v}i \quad \text{and} \quad M_w^{-1}M_z i = \frac{x-u}{v} + \frac{y}{v}i.$$

(b) Then write $M_w^{-1}M_z i = k_u e^{-r} i$, where $r = r(z,w)$, $k \in K = SO(2)$ are the geodesic polar coordinates for $M_w^{-1}M_z i$. And show that

$$2\cosh r(z,w) = \left[y^2+v^2+(x-u)^2\right]/(vy).$$

3.2.6 *The Heat Equation on H Using Helgason's Transform*

This section is modelled on the discussion in Gangolli [191, 192]. We want to consider the non-Euclidean analogue of the method that was used in Example 1.2.1 of Sect. 1.2 to find the fundamental solution of the Euclidean heat equation. We shall assume that $f(kz) = f(z)$ for all $k \in SO(2)$ and z in H. Then take the Helgason transform with respect to z of the PDE in Eq. (3.35). This gives

$$\mathcal{H}u_t = \mathcal{H}(\Delta u).$$

Integration by parts or Green's theorem says that

$$\frac{\partial}{\partial t}\mathcal{H}u = s(s-1)\mathcal{H}u.$$

Thus $\mathcal{H}u = \mathcal{H}f \cdot e^{s(s-1)t}$. Using Exercise 3.2.19, it follows that

$$u(z,t) = f * \mathcal{S}(e^{s(s-1)t})(z),$$

where $\mathcal{S}$ denotes the inverse transform Eq. (3.33):

$$\begin{aligned} G_t(e^{-r}i) &= \left[\mathcal{S}(e^{s(s-1)t}\right](e^{-r}i) \\ &= \frac{1}{4\pi}\int_{v\in\mathbb{R}} e^{-(v^2+\frac{1}{4})t}P_{-\frac{1}{2}+iv}(\cosh r)\, v\tanh(\pi v)\, dv. \end{aligned}$$

This is formula (3.37). Moreover, Exercise 3.2.22 shows that the **fundamental solution of the heat equation** or **density for the non-Euclidean normal distribution** is

$$G_t(e^{-r}i) = \frac{\sqrt{2}}{(4\pi t)^{3/2}e^{t/4}}\int_r^{\infty}\frac{be^{-b^2/4t}db}{\sqrt{\cosh b - \cosh r}}. \tag{3.38}$$

Therefore

$$G_t(z) > 0 \quad \text{for all } t > 0 \text{ and } z \in H. \tag{3.39}$$

Moreover,

$$G_t(z) \to \delta \quad \text{as } t \to 0+, \tag{3.40}$$

where $\delta =$ the Dirac delta distribution on H; i.e., $f * \delta =$ f. To see Eq. (3.40) note that by the Plancherel theorem for H we have

$$\|f * G_t - f\|_2^2 = \frac{1}{4\pi}\int\limits_{\substack{s=\frac{1}{2}+iv \\ v\in\mathbb{R}}} \left|(\widehat{f * G_t} - \hat{f})(s)\right|^2 v\tanh(\pi v)\, dv.$$

Now

$$\widehat{f * G_t} = \hat{f}\cdot\widehat{G_t} = \hat{f}\cdot e^{s(s-1)t} \to \hat{f} \quad \text{as } t \to 0+.$$

Thus,

$$\|f * G_t - f\|_2^2 \to 0 \quad \text{as } t \to 0+,$$

by the Lebesgue dominated convergence theorem.

In fact, G_t has the properties of the Gauss kernel, except that it does not appear to be possible to find a simpler formula for G_t than Eq. (3.37) or Eq. (3.38). Formula (3.38) was obtained by Karpelevich et al. [337], as well as the corresponding result for the three-dimensional analogue of H to be discussed in Volume II ([667], where the fundamental solution of the heat equation on hyperbolic 3-space is considered).

3.2.7 The Central Limit Theorem for K-Invariant Random Variables on H. Transmission Lines with Random Inhomogeneities

There are many versions of the central limit theorem on symmetric spaces such as the upper half-plane. The first papers on the subject were written by mathematicians in the U.S.S.R. during the 1950s and 1960s (see Karpelevich et al. [337], Tutubalin [688], and Virtser [705]). Mathematicians in the U.S. discussed these matters using very general limit theorems on stochastic differential equations in the 1970s (see Burridge and Papanicolaou [73], Keller and Papanicolaou [349], and Papanicolaou [509]). Mathematicians in France obtained central limit theorems for various sorts of Lie groups (see Bougerol [58], Clerc and Roynette [92], and Faraut [173]). Limit theorems have also been obtained by Heyer [294] and Trimèche [685]. A similar discussion of Euclidean rotation-invariant random variables is to be found in Kingman [352]. There are many potential applications for such limit theorems; e.g., in demography (see Cohen [96, especially p. 290]), learning theory (see LePage [412] and the references there), atomic physics (see Hantsch and von Waldenfels [248], as well as Hurt and Hermann [310]). Other related papers are Dudley [141, 142], Furstenberg [188], Furstenberg and Kesten [189], Gangolli [193], Getoor [209], and Letac [413]. The last reference gives analogues of classical probability problems for various symmetric spaces, including the p-adic symmetric spaces.

First, let us consider an engineering problem of Gertsenshtein and Vasil'ev [208]—a problem which requires the non-Euclidean central limit theorem for its solution. We wish to analyze a very long lossless transmission line with random inhomogeneities caused perhaps by tiny defects. Such inhomogeneities produce reflected waves, the properties of which are described by the reflection coefficient which we can view as a random variable Z in H, as we saw at the end of Sect. 3.1. The **composite** of two random inhomogeneities with reflection coefficients Z_1 and Z_2 in H produces a reflection coefficient

$$Z_1 \circ Z_2 = M_{Z_1} M_{Z_2} i, \quad \text{where } Z_j = M_{Z_j} i \text{ for } M_{Z_j} \in SL(2,\mathbb{R}), \ j = 1,2. \tag{3.41}$$

This composition is only well defined when the Z_j are $SO(2)$-invariant, which is the only case to be considered here. The question then arises as to the distribution of $Z_1 \circ \cdots \circ Z_n$, when it is correctly normalized, as $n \to \infty$. Having found the limit distribution, one should be able to compute the mean power output, for example.

To carry out this project, we need a central limit theorem for H. Luckily we already have a candidate density for the non-Euclidean normal distribution, namely the fundamental solution of the heat equation on H given by Eqs. (3.37) and (3.38). We shall attempt to keep our discussion as close as possible to the discussion of the Euclidean central limit theorem (see Theorem 1.2.7 of Sect. 1.2).

First, we should set down the requisite definitions. A **random variable** Z in H has **distribution function**

$$P(Z \in A) = \int_A f_Z(z) y^{-2}\, dx\, dy = \int_A f_Z\, d\mu,$$

where f_Z is the **density function** and

$$f_Z \geq 0, \ \int_H f_Z(z)\, d\mu = 1.$$

We shall consider only ***SO(2)*-invariant random variables** on H; this means that the density function must satisfy the invariance condition

$$f_Z(kz) = f_Z(z) \quad \text{for all } \ z \in H, \ k \in K.$$

The **sum** or **composition of two *SO(2)*-invariant random variables** is defined by formula (3.41).

We shall say that the random variables Z_1 and Z_2 are **independent** if

$$P(Z_1 \in A \text{ and } Z_2 \in A) = P(Z_1 \in A) P(Z_2 \in A).$$

Exercise 3.2.26. Show that if Z_1 and Z_2 are $SO(2)$-invariant independent random variables in H with density functions f_1 and f_2, respectively, then the density function for $Z_1 \circ Z_2$ is $f_1 * f_2$, with convolution defined as in Exercise 3.2.19.

Hint. *You can imitate the proof that works in the Euclidean case, since*

$$P(Z_1 \circ Z_2 \in A) = \int\limits_{M_{Z_2} M_{Z_1} i \in A} \int f_1(z_1) f_2(z_2)\, d\mu(z_1)\, d\mu(z_2) \quad \text{with } f_j = f_{Z_j}.$$

Set $w = M_{Z_2} M_{Z_1} i$, *and note that*

$$\int_{w \in A} f_{Z_1 \circ Z_2}(w)\, d\mu(w) = \int_{z_1 \in H} f_1(z_1) \int_{w \in A} f_2(M_W M_{Z_1}^{-1} i)\, d\mu(w)\, d\mu(z_1).$$

The **characteristic function** φ_Z of an $SO(2)$-invariant random variable Z in H is the **Helgason or non-Euclidean Fourier transform** from Eq. (3.29):

$$\varphi_Z(p) = \int_H f_Z(z) \left(\overline{\operatorname{Im} z}\right)^s d\mu(z), \quad s = \frac{1}{2} + ip, \ p \in \mathbb{R}. \tag{3.42}$$

These non-Euclidean characteristic functions possess many (but unfortunately not all) of the properties of Euclidean characteristic functions described in Theorem 1.2.6 of Sect. 1.2.

Exercise 3.2.27 (Properties of Characteristic Functions). Which properties of Euclidean characteristic functions in Theorem 1.2.6 of Sect. 1.2 fail for non-Euclidean characteristic functions?

Hint. *The main problem lies with part (2) of Theorem 1.2.6 of Sect. 1.2.*

In the preceding exercise it was noted that φ_Z lacks one important property of the Euclidean characteristic function. For φ_Z has domain $\mathbb{R}$ or $\mathbb{C}$ and not H itself. This means, for example, that we cannot say that if $a \in \mathbb{R}$, then $\varphi_{aZ}(p) = \varphi_Z(ap)$ for $p \in \mathbb{R}^+$. This property was quite important in our proof of the central limit theorem (Theorem 1.2.7 of Sect. 1.2). And the lack of this property seems to be the cause of some non-Euclidean trouble. Clerc and Roynette [92] meet this issue head on by considering random variables in the domain of the characteristic function rather than H. On the other hand, Karpelevich et al. [337] do not mention this problem and only sketch the beginnings of a proof of their central limit theorem. We shall combine ideas from both these papers.

The **density function for the non-Euclidean normal distribution** is defined to be $G_c(z)$, the fundamental solution of the non-Euclidean heat equation given by formulas (3.37) and (3.38). Then if N_c is a normally distributed random variable in H, the characteristic function is

$$\varphi_{N_c}(p) = \exp\left[-c\left(p^2 + \frac{1}{4}\right)\right].$$

In the non-Euclidean case, there are many possible analogues of the mean and the variance. We choose the most direct analogues and define for an $SO(2)$-invariant random variable Z in H, the **mean** m_Z and the **variance** or, better perhaps, the **second moment** d_Z by

$$\begin{aligned} m_Z &= 2\pi \int_{r>0} f_Z(e^{-r}i)\; r \sinh r\, dr, \\ d_Z &= 2\pi \int_{r>0} f_Z(e^{-r}i)\; r^2 \sinh r\, dr. \end{aligned} \tag{3.43}$$

Before proceeding with our discussion of the non-Euclidean central limit theorem, we need another asymptotic property of Legendre functions.

Exercise 3.2.28 (An Asymptotic Relation Between Legendre Functions and *J*-Bessel Functions).

(a) Show that

$$P_{-\frac{1}{2}+ip}(\cosh r) = \frac{\sqrt{2}}{\pi} \int_0^r \frac{\cos(pu)\, du}{\sqrt{\cosh r - \cosh u}}.$$

(b) Show that $P_{-\frac{1}{2}+ip}(\cosh r) \sim J_0(pr)$, as $r \to 0$.

The formula in part (b) of the preceding exercise is a special case of a very general phenomenon relating the spherical function on G/K to the spherical function on the tangent space to G/K at a point. We will consider the analogue for $SL(n,\mathbb{R})$ in Volume II [667].

The complete asymptotic expansion of $P_{-\frac{1}{2}+ip}(\cosh r)$, as $r \to 0$, can be found in Szegö [651] or Fock [181]. This has been generalized to rank 1 symmetric spaces by Stanton and Tomas [622]. Clerc and Roynette [92] use a similar result, which is analogous to

$$P_{-\frac{1}{2}+ipm}\left(\cosh \frac{r}{m}\right) \sim J_0(pr) \quad \text{as } m \to \infty.$$

In the classical case, the mean and standard deviation of a random variable are (essentially) the first and second derivatives of the characteristic function evaluated at 0. However, in the non-Euclidean case, this produces different integrals.

Exercise 3.2.29 (Other Analogues of the Mean and the Variance).

(a) Show that if Z is an $SO(2)$-invariant random variable in H, then

$$\varphi_Z'(0) = 0.$$

This can be viewed as a non-Euclidean analogue of the mean, which differs from that in Eq. (3.43).

(b) Define the **dispersion** D_Z of an $SO(2)$-invariant random variable in H by

$$D_Z = -\frac{\varphi_Z''(0)}{\varphi_Z(0)} = \frac{\displaystyle\int_{r \geq u \geq 0} \frac{u^2 f_Z(e^{-r}i)\sinh r}{\sqrt{\cosh r - \cosh u}}\, du\, dr}{\displaystyle\int_{r \geq u \geq 0} \frac{f_Z(e^{-r}i)\sinh r}{\sqrt{\cosh r - \cosh u}}\, du\, dr}.$$

Hint. *Use Exercise 3.2.28(a).*

Note that D_Z can be viewed as a non-Euclidean analogue of the variance, which differs from that in Eq. (3.43).

(c) Show that if Z_1 and Z_2 are independent $SO(2)$-invariant random variables in H, then

$$D_{Z_1 \circ Z_2} = D_{Z_1} + D_{Z_2}.$$

(d) Show that $D_{N_c} = 2c$, if N_c has the non-Euclidean Gaussian or normal distribution; i.e., has density given by the heat kernel G_c for H.

Exercise 3.2.30. Graph the heat kernel $G_c(e^{-r}i)$ for various values of c. Compare with Fig. 3.13. Then compute the mean and second moment defined by

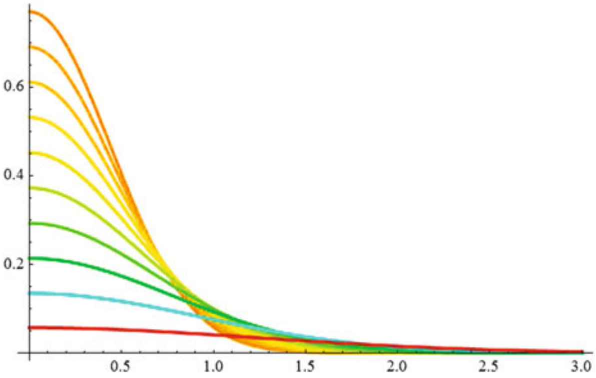


Fig. 3.13 Non-Euclidean heat kernel $G_c(e^{-r}i)$ for $c = 1/n,\ \ n = 1,2,\dots,10$

formula (3.43) for these same values of c. Show that the mean approaches infinity as c approaches infinity. This can be derived from the following formula of Gertsenshtein and Vasil'ev [208]:

$$\int_1^\infty G_c(e^{-r}i)r\,dr = e^{2c}.$$

To discuss the non-Euclidean central limit theorem, we suppose that we are given a sequence $\{Z_n\}_{n\geq 1}$ of independent, $SO(2)$-invariant, random variables in H, each having the same density function f. We want to find some way to normalize the random variable

$$S_n = Z_1 \circ \cdots \circ Z_n \tag{3.44}$$

in order to be able to say that the normalized variable, which we shall call $S_n^\#$, approaches the random variable with density $G_c =$ the fundamental solution of the non-Euclidean heat equation given by formulas (3.37) or (3.38) above. Now Karpelevich et al. [337] normalize S_n by noting for $A \subset H$,

$$P(S_n \in A) = 2\pi \int_{e^{-r}i\in A, r>0} f_{S_n}(e^{-r}i)\ \sinh r\,dr. \tag{3.45}$$

Thus we really have densities on $\mathbb{R}^+$, using the non-Euclidean radial variable ($r =$ non-Euclidean distance of $z \in H$ to $i = \sqrt{-1}$). We want to normalize S_n by dividing the radial coordinate random variable by $\sqrt{n}$. Thus the **characteristic function of the normalized random variable** $S_n^\#$ is

$$\varphi_{S_n^\#}(p) = 2\pi \int_{r>0} f_{S_n}(e^{-r}i)\sinh r\,P_{-\frac{1}{2}+ip}(\cosh(r/\sqrt{n}))\,dr, \tag{3.46}$$

In the classical case, one could easily move the $n^{-1/2}$ over to the p-variable, since if X is a random variable in $\mathbb{R}$,

$$\varphi_{X/\sqrt{n}}(p) = \int_{x\in\mathbb{R}} f(x)\exp(ipx/\sqrt{n})dx = \varphi_X(p/\sqrt{n}).$$

But the relation between $P_{-1/2+ip/\sqrt{n}}(\cosh r)$ and $P_{-1/2+ip}(\cosh(r/\sqrt{n}))$ appears to be rather bizarre at first sight.

However, we can use Exercise 3.2.28(b) to rectify this situation and find that

$$\varphi_{S_n^\#}(p) \sim \left(2\pi\int_{r>0} f(e^{-r}i)\ \sinh r\, J_0\left(\frac{rp}{\sqrt{n}}\right) dr\right)^n \quad \text{as } n\to\infty.$$

And we have the power series for the J-Bessel function:

$$J_0(x) = \sum_{k\geq 0}\frac{(-1)^k}{(k!)^2}\left(\frac{x}{2}\right)^{2k}.$$

This implies that

$$\begin{aligned}\varphi_{S_n^\#}(p) &\sim \left(2\pi\int_{r>0} f(e^{-r}i)\sinh r\, dr - \frac{2\pi}{4n}p^2\int_{r>0} f(e^{-r}i)r^2\sinh r\, dr\right)^n\\ &\sim \left(1-\frac{dp^2}{4n}\right)^n \sim e^{-dp^2/4} \quad \text{as } n\to\infty,\\ d &= d_{Z_n} = 2\pi\int_{r>0} f(e^{-r}i)r^2\sinh r\, dr,\end{aligned}$$

as in formula (3.43).

In order to complete the discussion of the central limit theorem for rotation-invariant random variables on H, one must imitate the argument given in the proof of the Euclidean central limit theorem (Theorem 1.2.7 of Sect. 1.2). We have proved that

$$\lim_{n\to\infty}\varphi_{S_n^\#}(p) = \exp(-dp^2/4). \tag{3.47}$$

Note that the density function for the random variable $S_n^\#$ is

$$f_n^\#(ke^{-r}i) = \sqrt{n}(f*\cdots*f)(e^{-r\sqrt{n}}i)\sinh(r\sqrt{n})/\sinh r, \tag{3.48}$$

where $k\in SO(2)$ and $r>0$. Next let $\alpha\in C_c^\infty(H)$. And set

$$d\sigma(p) = \frac{1}{4\pi}p\tanh(\pi p)\, dp.$$

By formula (3.47), the Plancherel theorem (Theorem 3.2.3), and the Lebesgue dominated convergence theorem, we have

$$\begin{aligned}\lim_{n\to\infty}\int_H f_n^{\#}(z)\alpha(z)\,d\mu &= \lim_{n\to\infty}\int_{-\infty}^{\infty}\mathcal{H}f_n^{\#}\left(\frac{1}{2}+ip\right)\mathcal{H}\alpha\left(\frac{1}{2}+ip\right)\,d\sigma(p)\\ &= \int_{-\infty}^{\infty}\exp(-dp^2/4)\mathcal{H}\alpha\left(\frac{1}{2}+ip\right)\,d\sigma(p) \qquad (3.49)\\ &= e^{d/4}\int_H G_{d/4}(z)\alpha(z)\,d\mu.\end{aligned}$$

We can approximate the indicator function of a measurable set in H by α in $C_c^\infty(H)$ to complete the proof of the following theorem.

Theorem 3.2.4 (A Non-Euclidean Central Limit Theorem for Rotation-Invariant Random Variables). *Suppose that $\{Z_n\}_{n\geq 1}$ is a sequence of independent, $SO(2)$-invariant random variables in H, each having the same density function $f(z)$. Let $S_n = Z_1\circ\cdots\circ Z_n$ be normalized as in formulas (3.46) and (3.48) above. Suppose that $f_n^{\#}$ is the density function for the normalized random variable $S_n^{\#}$. Then for measurable sets $A\subset H$ we have*

$$\int_A f_n^{\#}(z)d\mu \sim e^{d/4}\int_A G_{d/4}(z)d\mu \quad \text{as } n\to\infty.$$

Here G_c denotes the fundamental solution of the non-Euclidean heat equation or density for the normal distribution given by formulas (3.37) and (3.38). And d is defined by formula (3.43).

Exercise 3.2.31.

(a) Prove that $|P_s(\cosh r)|\leq 1$, if $-1\leq \operatorname{Re} s\leq 0$.
(b) Use (a) to show that if φ_Z is the characteristic function of an $SO(2)$-invariant random variable Z with density function $f(z)$, then $|\varphi_Z(p)|\leq 1$ for all $p\in\mathbb{R}$.
(c) Justify formula (3.49).

Finally, let us return to the discussion of the lossless transmission line with random inhomogeneities. Exercise 3.2.30 showed that the mean distance from i to a normally distributed random variable with density $G_c(z)$ increases exponentially as c approaches infinity. And we can conclude from parts (c) and (d) of Exercise 3.2.29 that c approaches infinity as the length of the transmission line increases. Recalling the discussion in Sect. 3.1 which showed that the transmitted power decreases as the distance of the reflection coefficient from the origin increases (measuring distance using the non-Euclidean metric), we conclude that a long transmission line reflects almost all of the incoming power. More precise calculations might allow an engineer to do something about this (see Feller [177, Vol. II, pp. 208–209]).

There are higher rank versions of the central limit theorem. These will be discussed in Volume II [667]. The reader should compare the discussion of the central limit theorem for $\mathcal{P}_3$ in Vol. II [667, pp. 109–110] with the preceding discussion. See also Richards [541].

3.2.8 *Some Remarks on the Uses in Theoretical Physics of Harmonic Analysis on Groups Such as* $SL(2,\mathbb{R})$

We shall keep our remarks very brief. Detailed discussions of the representations of $SL(2,\mathbb{R})$ can be found in Lang [389], Sugiura [648], and Vilenkin [704]. Applications of representations of noncompact groups like $SL(2,\mathbb{R})$ to physics are considered, for example, in Barut and Raçzka [28], Mackey [442–444], and Wybourne [746].

Physicists are interested in representations of groups such as $SO(3,1)$ and $SO(4,2)$ which leave invariant various PDEs describing, for example, electromagnetic waves or elementary particles. The representations of groups such as $SO(p,q)$ have seen much attention. These are the real $n = p+q$ by n matrices which leave invariant the quadratic form $x_1^2 + \cdots + x_p^2 - x_{p+1}^2 \cdots - x_n^2$. We saw in Exercise 3.1.13 of Sect. 3.1 that $SL(2,\mathbb{R})$ and $SO(2,1)$ are closely related. In fact, they have the same Lie algebra. See Borel and Wallach [55], Knapp [356], or Vogan [707] for more information on representations of Lie groups such as $SL(n,\mathbb{R})$ or $GL(n,\mathbb{R})$ or $SO(p,q)$.

The Fourier transform of a function on $G = SL(2,\mathbb{R})$ involves various series of representations of G such as the principal continuous series and the discrete series (see Michelle Vergne [697] or Lang [389]). And Fourier analysis on $SL(2,\mathbb{R})$ involves a mixture of Fourier series and integrals. The inversion formula goes back to Bargmann [25] and Harish-Chandra [252]. We shall not state it here. Instead, let us just say a few words about the physical relevance of such expansions. A good introduction to this subject is Chap. 18 of Wybourne [746], where it is shown that many physics problems lead to the Lie algebra of $SL(2,\mathbb{R})$. In particular, one can find the Lie algebra of $SL(2,\mathbb{R})$ or equivalently, $SU(1,1)$, in the ODE:

$$f''(y) + q(y)f(y) = 0, \quad q(y) = ay^{-2} + by^2 + c.$$

In fact, this sort of argument allows one to being $SO(2,1)$ into the discussion of the hydrogen atom. Then one can bring in $SO(4)$ and $SO(4,2) \cong SU(2,2)/\{\pm I\}$. In this way more accurate information on the spectrum of hydrogen has been obtained.

Clearly, solutions of non-Euclidean wave equations and their properties such as the truth or falsity of the Huygens' principle are of interest in general relativity (see Friedlander [186] and Helgason [276]).

3.3 Fundamental Domains for Discrete Subgroups Γ of $G = SL(2,\mathbb{R})$

> A two-dimensional smooth orientable, but not compact space of constant negative curvature with the topology of a torus [or a sphere] is investigated. It contains an open end; i.e., an exceptional point at infinite distance, through which a particle or wave can enter or leave, as in the exponential horn of certain antennas or loud-speakers.
>
> —From Gutzwiller [239, p. 341]. See Exercise 3.3.14 for more details.

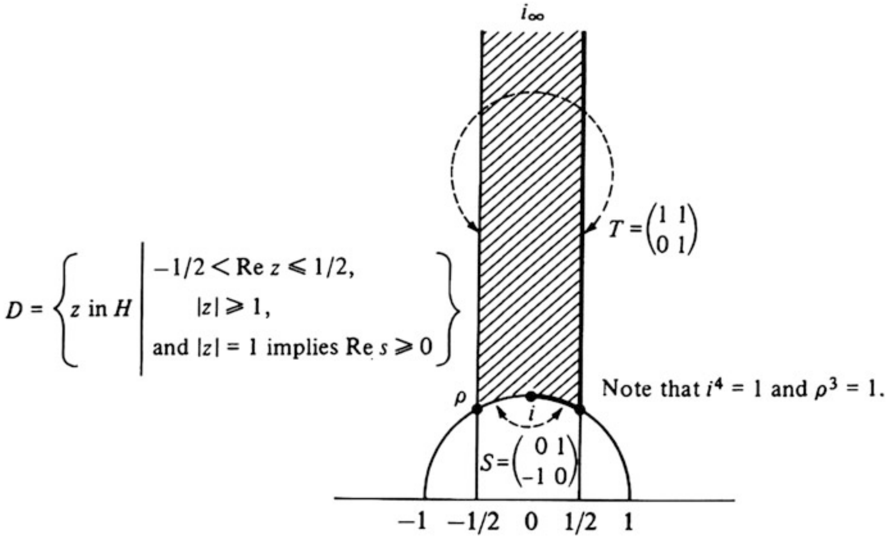

$i\infty$
$T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$
$D = \left\{ z \text{ in } H \;\middle|\; \begin{array}{l} -1/2 < \operatorname{Re} z \leqslant 1/2, \\ |z| \geqslant 1, \\ \text{and } |z| = 1 \text{ implies } \operatorname{Re} s \geqslant 0 \end{array} \right\}$
Note that $i^4 = 1$ and $\rho^3 = 1$.
ρ
i
$S = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$
-1 $-1/2$ 0 $1/2$ 1

Fig. 3.14 A non-Euclidean triangle D through the points $\rho = e^{2\pi i/3}$, $\rho + 1$, $i\infty$, which is a fundamental domain for H mod $SL(2,\mathbb{Z})$. The domain D is *shaded*. *Arrows* show boundary identifications by the fractional linear transformations from S and T which generate $SL(2,\mathbb{Z})/\{\pm 1\}$

3.3.1 Introduction

We saw in Sect. 3.2 that it was quite useful to study $\mathbb{R}/\mathbb{Z}$, which we think of as the circle or the interval $[0,1]$ with 0 and 1 identified. And in Sect. 3.4, we saw that it is valuable to study crystals $M(3,\mathbb{R})/\Gamma$, where $M(3,\mathbb{R})$ is the Euclidean group and Γ is a discrete subgroup of $M(3,\mathbb{R})$. Similarly, there are higher-dimensional crystals $M(n,\mathbb{R})/\Gamma$, with Γ any discrete subgroup of the n-dimensional Euclidean motion group $M(n,\mathbb{R})$, which is a semidirect product of $O(n)$ and the translation group $\mathbb{R}^n$. Bieberbach proved in 1910 that $M(n,\mathbb{R})$ has only a finite number of different subgroups with compact fundamental domains. This answered part of the 18th of Hilbert's famous problems (see Milnor [468, pp. 491–497]).

In this section, we shall consider discrete subgroups Γ of $SL(2,\mathbb{R})$. It was proved by Poincaré that, unlike $\mathbb{R}^n$, the upper half-plane has an infinite number of essentially different kinds of compact fundamental domains $\Gamma\backslash H$. Thus Hilbert called these domains Poincaré polygons. We shall be most interested, however, in the noncompact fundamental domain for the **modular group** $\Gamma = SL(2,\mathbb{Z})$ of 2×2 integer matrices of determinant 1. A fundamental domain for this group is pictured in Fig. 3.14. The main goal of this chapter is to study harmonic analysis on $SL(2,\mathbb{Z})\backslash H$ (see Sect. 3.7). In Volume II [667] we aim to discuss the analogue for $SL(n,\mathbb{Z})$ and other more complicated discrete groups.

General references for this section are Apostol [12], Ford [183], Gunning [236], Hecke [258, 259], Svetlana Katok [342], Klein and Fricke [355], Knopp [357],

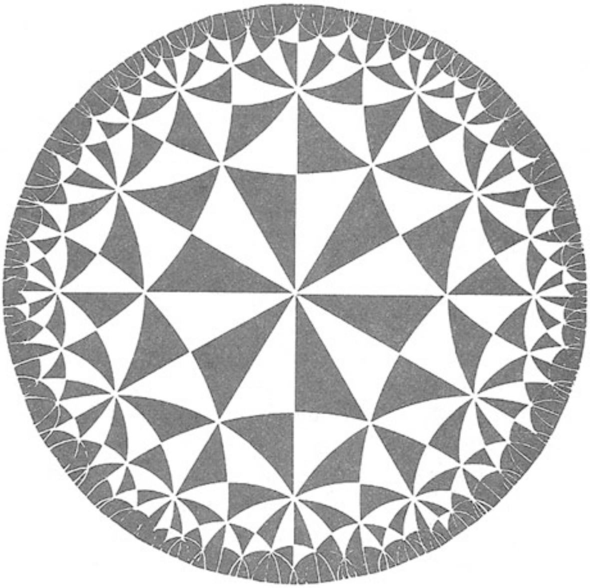


Fig. 3.15 The Coxeter illustration that inspired Escher. (From Coxeter [114, p. 285]. Reprinted by permission of John Wiley)

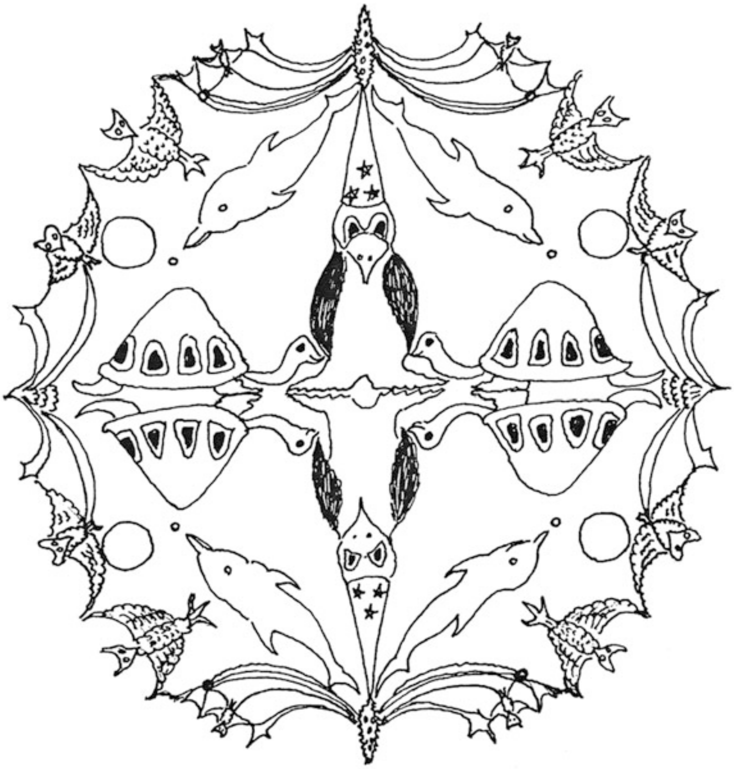
Koblitz [359], Lang [391], Lehner [410], Maass [437], Rankin [532], Schoeneberg [562], Serre [576], Shimura [589], and Siegel [596].

Just as one can tile the Euclidean plane with L-translates of fundamental domains $\mathbb{R}^2/L$ for two-dimensional lattices L, one can tile H with Γ-translates of a fundamental domain $\Gamma\backslash H$. This produces what is called a non-Euclidean tiling or tessellation of H and an interesting picture, sometimes referred to as non-Euclidean wallpaper. In 1958, the artist M. C. Escher saw such a picture in a book of H. S. M. Coxeter (Fig. 3.15) and was inspired to create his various circle limit pictures (see Ernst [166, p. 108]). In Escher's circle limit pictures, the upper half-plane is replaced by the unit disc, using the Cayley transform, from formula (3.10) of Sect. 3.1. Figure 3.16 is the tessellation of H for $\Gamma = SL(2,\mathbb{Z})$. It was drawn by computer (as directed by Mark Eggert). In Volume II [667], we will find some beautiful three-dimensional analogues of this tessellation, also drawn by computer with Mr. Eggert's direction. You can find other two-dimensional tessellations in Ford [183, pp. 305–309], Hurwitz and Courant [312, pp. 430–445], Klein and Fricke [355], Lehner [410, pp. 11–17], and Nehari [492, pp. 308–316]. See also the websites of David Joyce

http://aleph0.clarku.edu/~djoyce/poincare/poincare.html,

Helena Verrill

https://www.math.lsu.edu/~verrill/fundomain/,



and Gerard Westendorp

http://westy31.home.xs4all.nl/Geometry/Geometry.html.

Why study harmonic analysis on $\Gamma\backslash H \cong \Gamma\backslash G/K,\ G = SL(2,\mathbb{R}),\ K = SO(2)$, and the analogue for other Lie groups? One justification comes from the following dialogue which appeared in 1953 in *Scientific American* (see Le Corbeiller [403]):

> "In the last 60 years, however, a new revolution has taken place, and everywhere we look we find that what seems to be continuous is really composed of atoms...."
>
> "But are not modern mathematicians interested in such things?" asked Empeiros [a physicist].
>
> "They are," I answered, "but they give them other names. They call them Number Theory and the Theory of Discontinuous Groups. Actually they have found much more than we can use as yet in physics, but we have in crystals illustrations of some of their simpler theorems...."

Functions on $\Gamma\backslash H$ (or equivalently, on the fundamental domain with boundaries identified) are called **automorphic functions**. Surprisingly, **every analytic function *f(z)* can be considered as an automorphic function**. For f lives on a Riemann surface S which has a (simply connected) universal covering surface $\hat{S}$. The Riemann mapping theorem identifies $\hat{S}$ with either the Riemann sphere, the complex plane or H. The group of isometries of S contains a subgroup Γ isomorphic to the group of covering transformations of $\hat{S}$ over S. Thus f is automorphic with respect to Γ. For the details of this argument which we first saw in a course by Joseph Lehner, see Siegel [596, Vol. I, Chap. 2], or Hurwitz and Courant [312, pp. 453–550].

Historically, the study of automorphic functions arose with the attempt to solve various problems arising in applied mathematics, as well as number theory. For example, the problem of heat diffusion leads to **theta functions** (see Exercise 1.3.9 of Sect. 1.3). The search for a proof of the prime number theorem, giving the asymptotic behavior of the number of primes less than or equal to x as $x \to \infty$, leads to the study of Riemann's zeta function. The latter function is connected with the theta function by Mellin transform (see Exercise 1.4.7 of Sect. 1.4). The simplest theta function is

$$\theta(z) = \sum_{n\in\mathbb{Z}} \exp(\pi i n^2 z) \quad \text{for } z \in H. \tag{3.50}$$

Theta is not quite an automorphic function. It is what is known as an **automorphic form**, since theta has the following transformation properties:

$$\begin{aligned} \theta(z+2) &= \theta(z), \\ \theta(-1/z) &= (z/i)^{1/2}\theta(z). \end{aligned}$$

The last equation results from Poisson summation (see Exercise 1.3.8 of Sect. 1.3 and Exercise 1.4.6 of Sect. 1.4). The two maps $z \mapsto z+2$ and $z \mapsto -1/z$ generate a discrete subgroup of $SL(2,\mathbb{R})/\{\pm I\}$ called the **theta group** Γ_θ. Also, $\theta(2z)$ is a modular form of weight $\frac{1}{2}$ for the congruence group $\Gamma_0(4)$ (see Pfetzer [513] and Serre and Stark [581]). The **congruence subgroup** $\Gamma_0(N)$ of $SL(2,\mathbb{Z})$ is defined by

$$\Gamma_0(N) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \,\middle|\, a,b,c,d \text{ in } \mathbb{Z},\ ad-bc=1,\ c \equiv 0(\text{mod} N) \right\}.$$

Many problems of mathematics and physics lead to **elliptic integrals,** for example the lemniscate integral:

$$w = f(z) = \int (1-z^4)^{-1/2}\, dz.$$

The **Dirichlet problem** for various domains in the plane can often by solved via such integrals, thanks to the Schwarz–Christoffel transformation (see Carrier et al. [76, Chap. 4]). It turns out that the inverse function $z = g(w)$ is doubly periodic (see Ahlfors [3, p. 232]). The theory of such functions leads one quickly to automorphic functions and forms. In fact, one is often able to compute elliptic integrals via theta functions (see Sect. 3.4). Other references for these things are Lehner [410], Nehari [492, pp. 280–296, 308–316], and Siegel [596, Vols. I, II].

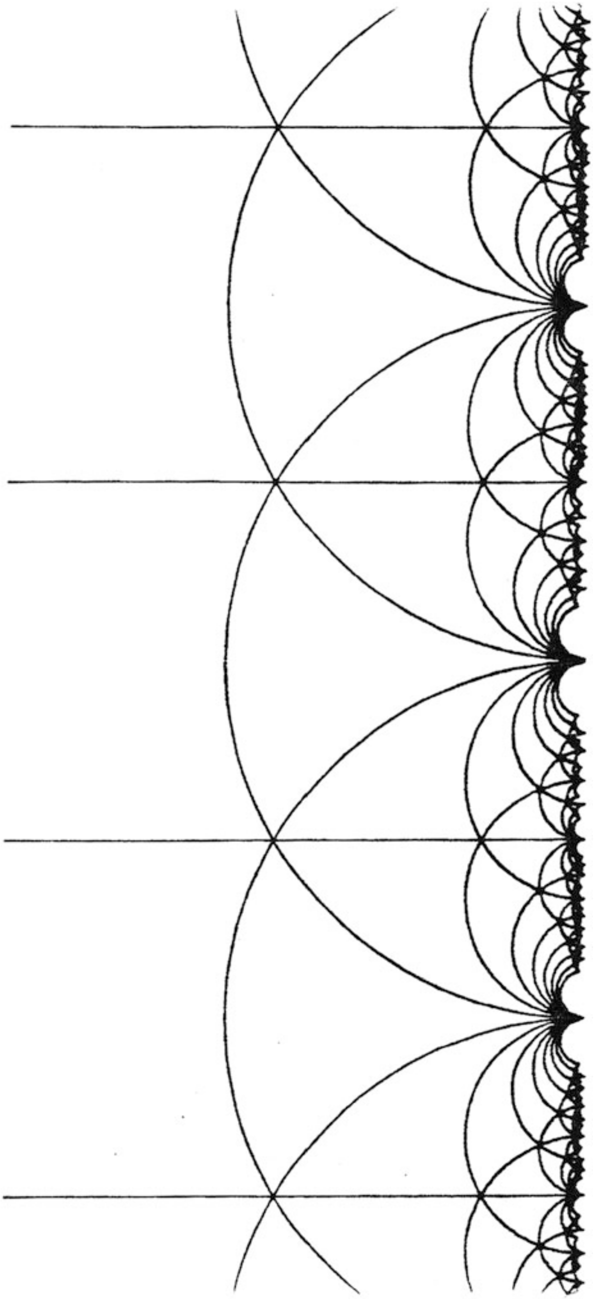


Fig. 3.16 Tessellation of H for $SL(2,\mathbb{Z})$. (Computer drawing by the UCSD VAX and Mark Eggert)

The J-function is another example of an automorphic or modular function that can be viewed as giving a certain conformal mapping. It can also be used to prove **Picard's theorem** that an entire function takes on every finite value with one possible exception (see Sect. 3.4). The harder Picard theorem says that a function holomorphic in a punctured disc with an essential singularity at the center has the property that there is at most one finite number a such that the function takes on the value a only finitely many times. There is also a way of dealing with the harder result via automorphic forms (see Lehner [411]).

There are many applications of automorphic forms to algebra and number theory. The computation of a fundamental domain for $SL(2,\mathbb{Z})\backslash H$ yields an **algorithm for the computation of class numbers of imaginary quadratic fields** (see Exercise 3.3.5, which follows). This is just the classical correspondence between ideal classes of imaginary quadratic fields and reduced integral positive binary quadratic forms (see Borevitch and Shafarevitch [56, pp. 149 ff], or Davenport [120, p. 195 ff]). The same fundamental domain can also be considered a fundamental domain for the **reduction theory of positive definite quadratic forms of determinant 1** by Exercise 3.1.9 of Sect. 3.1 and Siegel [597, pp. 68–74].

Hermite, Kronecker, and Brioschi used automorphic functions to **solve the general algebraic equation of degree 5** (see Lehner [410, p. 10]). Automorphic functions are also necessary for **explicit versions of class field theory**—the study of normal extensions of algebraic number fields with abelian Galois group. The J function is useful here, since it has the property that $J(a)$ is an algebraic integer when a lies in the upper half-plane and an imaginary quadratic field. More details on this subject can be found in Borel and Chowla [53]. **Hilbert's problem 12** asks for functions whose special values generate abelian extensions of number fields, just as values of the exponential function generate abelian extensions of the field $\mathbb{Q}$ of rational numbers. The answer for imaginary quadratic base fields requires automorphic functions and was found in the late 1800s and early 1900s by Weber, Kronecker, Fueter, and others. Much of the work on the subject is due to G. Shimura (see [589]). See also Stark [626, 627].

Recently automorphic forms have appeared in the proof of **Fermat's Last Theorem** by Wiles and Taylor. See Mozzochi [480] for a popular account of the proof. We will say more about this later as well as many more number theoretic applications of automorphic forms.

The study of the **representations of** ***SL(2,*** $\mathbb{R}$ ***)*** also leads to classical automorphic forms in the case of discrete series representations (see Gelfand et al. [203, pp. 43–48]). So far the only functions on H that we have mentioned in this introduction are complex analytic or meromorphic functions of z in H (see Sect. 3.4). However, there are other interesting automorphic functions. Their study is more recent. They correspond to the continuous series representations of $SL(2,\mathbb{R})$ and they are eigenfunctions of the non-Euclidean Laplacian invariant under $SL(2,\mathbb{Z})$. Such functions were first studied systematically in 1949 by Maass [439]. We will consider these Maass waveforms in Sect. 3.5. They are essential for harmonic analysis on $\Gamma\backslash H$ to be discussed in Sect. 3.7.

Automorphic forms and harmonic analysis on fundamental domains of discrete subgroups of Lie groups have appeared in many papers in physics journals (see Bogomolny [46], Bolte and Steiner [47], Gutzwiller [239–241] and Hurt [308,309]). For example, Monastyrsky and Perelomov [474] consider automorphic forms on general symmetric spaces and give several references to applications, one of which concerns the problem of boson pair creation in alternating external fields. These applications are basically a reflection of the usefulness of group representations in physics as we mentioned at the end of Sect. 3.2, plus the fact that automorphic forms are involved in the description of various series of representations of the Lie groups like $SL(2,\mathbb{R})$. The terminology used in the quantum mechanics work is that the automorphic forms appear in **systems of coherent states**.

3.3.2 *Fundamental Domain for* $SL(2,\mathbb{Z})$

After this, perhaps overlong, introduction to our subject, let us get down to the study of $\Gamma\backslash H$, for discrete subgroups Γ of $SL(2,\mathbb{R})$. For concreteness, we shall mostly restrict our study to the modular group $\Gamma = SL(2,\mathbb{Z})$. A **fundamental domain** D for Γ is a nice subset D of H which behaves like the quotient space $\Gamma\backslash H$ (at least up to boundary identifications). Thus for every z in H, there is a γ in Γ with γz in D. Moreover, if z and w lie in the interior of D and $z = \gamma w$, then γ is either the identity matrix I or $-I$. We shall say that z and γz are **equivalent under** Γ if $\gamma \in \Gamma$. As in Sect. 3.1, the action of $\gamma \in \Gamma$ on z in H is by fractional linear transformation. The example to keep in mind is the interval $[0,1)$ as a fundamental domain for $\mathbb{R}/\mathbb{Z}$.

Exercise 3.3.1 (A Fundamental Domain for SL(2, $\mathbb{Z}$)). Show that a fundamental domain for $SL(2,\mathbb{Z})$ can be taken as the region pictured in Fig. 3.14; that is,

$$D = \left\{ z \in H \;\middle|\; -\frac{1}{2} < \operatorname{Re} z \leq \frac{1}{2},\ |z| \geq 1,\ \text{and if}\ \ |z| = 1, \text{then}\ \ \operatorname{Re} z \geq 0 \right\}.$$

Hints.

(1) **The Highest Point Method.** Suppose $z \in H$. You must find γ in $\Gamma = SL(2,\mathbb{Z})$ so that $\gamma z \in D$. Clearly you can translate z to make $\operatorname{Re}(\gamma z) \in \left(-\frac{1}{2}, +\frac{1}{2}\right]$. Next note that $\operatorname{Im}(\gamma z) = y|cz+d|^{-2}$, if $\gamma z = (az+b)/(cz+d)$, with $a,b,c,d \in \mathbb{Z}$ and $ad - bc = 1$. Now you can choose γ so that $|cz+d|$ is minimal; i.e., $\operatorname{Im}(\gamma z)$ is maximal, so that γz is a "highest point" equivalent to z. Why? To prove that $|z| \geq 1$, use the existence of $Sz = -1/z$ to obtain a contradiction to the maximality of $\operatorname{Im}(\gamma z)$ otherwise.

Next you must prove that $z, \gamma z \in D$ for $\gamma \in \Gamma$, implies that $\gamma = \pm I$. This can be done by brute force. Suppose $\gamma z = (az+b)/(cz+d)$. Then

$$|cz+d|^2 = c^2 z\bar{z} + cd(z+\bar{z}) + d^2 \geq (|c| - |d|)^2 + |cd|\,.$$

Therefore $c \neq 0$ implies that $|cz+d| \geq 1$. It follows that $\operatorname{Im} z = \operatorname{Im} \gamma z$ if $c \neq 0$, since you can also consider γ^{-1}. Then $(|c| - |d|)^2 + |cd| = 1$ and only a few possible matrices γ need to be considered: S, STS, $ST^{-1}S$, with S, T as in

Fig. 3.14. It is easy to see what these matrices do to the fundamental domain (see Fig. 3.17). The case $c = 0$ is even easier.

(2) **The Generators and Relations Method of Poincaré.** If you compute the tessellation of H by $SL(2,\mathbb{Z})$ as in Fig. 3.17, you may notice, as Poincaré did, that the fundamental domain gives generators and defining relations of the **projective linear group** $PSL(2,\mathbb{Z}) = SL(2,\mathbb{Z})/\{+I, -I\}$. The generators are the maps identifying the sides of the fundamental region. To have pairs of sides you need four sides, and not three. So cut the bottom circle in half at the point $i = \sqrt{-1}$. Then the maps identifying the sides of D in pairs are $\overline{T}z = z + 1$ and $\overline{S}z = -1/z$. Here we write $\overline{T}$ and $\overline{S}$ to denote the fractional linear transformations corresponding to the matrices T and S. This notation is easily confused with complex conjugate and I will usually try to avoid it.

The defining relations for $PSL(2,\mathbb{Z})$ are obtained by picture. To do this, follow the mappings that circle the two vertices $\rho = (-1 + \sqrt{-3})/2$ and $i = \sqrt{-1}$. Since these mappings must lead to the identity when composed, you find that $(\overline{ST})^3 = \overline{I}$ when you follow the copies of D around the vertex ρ and you find that $\overline{S}^2 = \overline{I}$ when you follow the copies of D around the point i.

Further Notes on the Preceding Exercise. Poincaré showed that one can start with a geodesic polygon P with certain properties and obtain a discrete group Γ for which the polygon is a fundamental domain. The sides of P must be geodesics arranged in pairs equivalent under certain fractional linear maps and equal in non-Euclidean length. The sums of angles at equivalent vertices in the fundamental domain which are fixed by order k elliptic fractional linear transformations must be $2\pi/k$. See the end of this section for the definition of “elliptic.” Thus, in the case of the modular group, the sum of the angles of the vertices of the fundamental domain D at ρ and $\rho + 1$ is $2\pi/3$ and 3 is just the order of the fractional linear map fixing ρ. And the sum of the angles at i is just $\pi = 2\pi/2$, and 2 is again the order of the fractional linear map fixing i. In order to complete the preceding exercise using Poincaré’s ideas you must show that the generators and relations theorem actually gives an algorithm for finding fundamental domains.

Stark [628] has shown how to generalize these considerations to higher-dimensional cases in order to obtain analogues of Poincaré’s results and to compute the fundamental domains. Other references for Poincaré’s theory are Beardon [31], Svetlana Katok [342], Lehner [410], and Poincaré [517].

Exercise 3.3.2.

(a) Use the Euclidean algorithm to show that $PSL(2,\mathbb{Z}) = SL(2,\mathbb{Z})/\{+1, -I\}$ is generated by $Sz = -1/z$ and $Tz = z + 1$.

Hint. *Note that*

$$\begin{pmatrix} a' & b' \\ c' & d' \end{pmatrix} = \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} a+tc & * \\ * & * \end{pmatrix}.$$

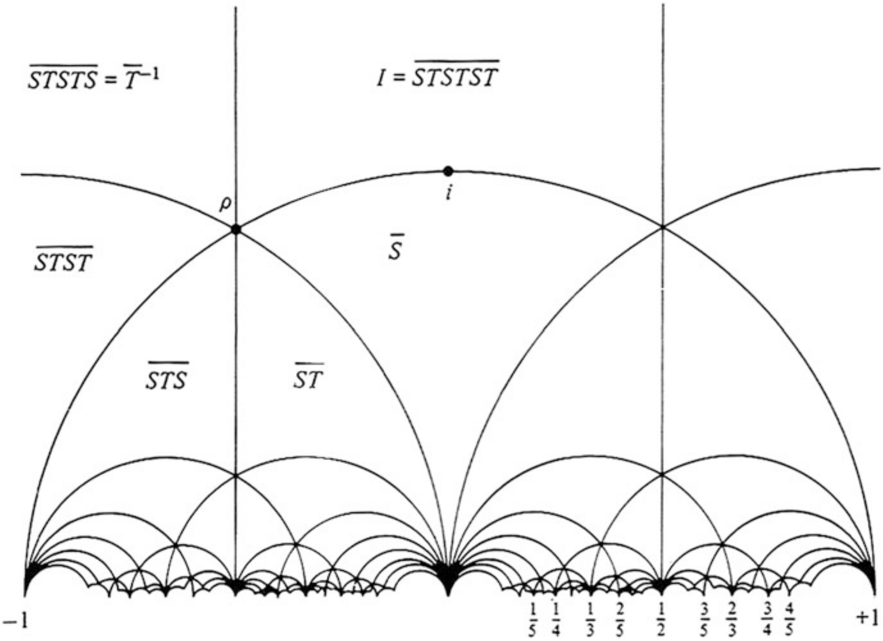

$\overline{STSTS} = \overline{T}^{-1}$
$I = \overline{STSTST}$
ρ
i
$\overline{S}$
$\overline{STST}$
$\overline{STS}$
$\overline{ST}$
-1
$\frac{1}{5}$ $\frac{1}{4}$ $\frac{1}{3}$ $\frac{2}{5}$ $\frac{1}{2}$ $\frac{3}{5}$ $\frac{2}{3}$ $\frac{3}{4}$ $\frac{4}{5}$
$+1$

Fig. 3.17 Poincare's method of generators and relations illustrated for $SL(2,\mathbb{Z})/\{+I,-I\}$. $\bar{S}z = -1/z$; $\overline{T}z = z+1$; $\rho = (-1+\sqrt{-3})/2$; $i = \sqrt{-1}$. Generators of $SL(2,\mathbb{Z})/\{+I,-I\}$ are $\overline{S}$ and $\bar{T}$. Defining relations are $\bar{S}^2 = \bar{I}$ and $(\overline{ST})^3 = \bar{I}$. Here we write $\overline{T}$ and $\overline{S}$ to denote the fractional linear transformations corresponding to the matrices T and S. The bars here do not refer to complex conjugate

> *You can choose $t \in \mathbb{Z}$ to make $0 \le a' < c$ by the Euclidean algorithm. Then you can use S to flip a and c. Finally, you can obtain $c = 0$ in a finite number of steps, since a and c are relatively prime.*

(b) Show that the algorithm in (a) also gives the continued fraction expansion of a/c (see Hardy and Wright [251]).

Exercise 3.3.3. Which of Figs. 3.18, 3.19, or 3.20 are tessellations of the unit disc coming from the modular group after mapping H to the unit disc under the Cayley transform $z \mapsto i(z-i)/(z+i)$? Why? In Figs. 3.18 and 3.19, a shaded and a light region together make up a fundamental domain.

Exercise 3.3.4. The fundamental domain for $SL(2,\mathbb{Z})$ in Fig. 3.14 is a non-Euclidean triangle with angles $\pi/3$ at $\rho = (-1+\sqrt{-3})/2$ and $\rho+1$, and angle 0 at ∞. Show that the non-Euclidean area of the fundamental domain for $SL(2,\mathbb{Z})$ is $\pi/3$. Does this agree with Exercise 3.1.7 of Sect. 3.1?

The fundamental domain for $SL(2,\mathbb{Z})$ was used by Gauss in his research on quadratic forms. Instead of quadratic forms, one can consider ideal classes in

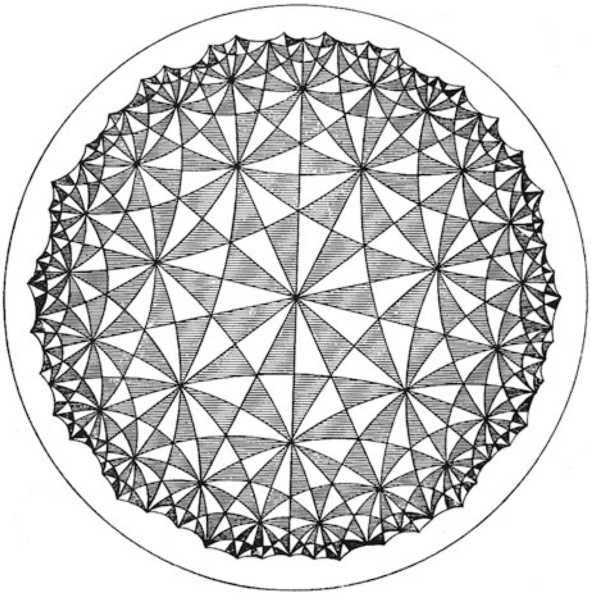


Fig. 3.18 Tessellation of the unit disc. (From Klein and Fricke [355]. Reprinted by permission of Teubner)

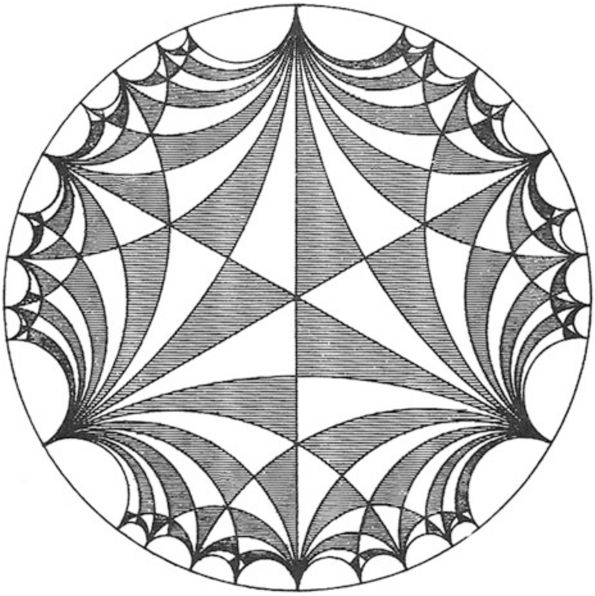


Fig. 3.19 Another tessellation of the unit disc. (From Klein and Fricke [355]. Reprinted by permission of Teubner)

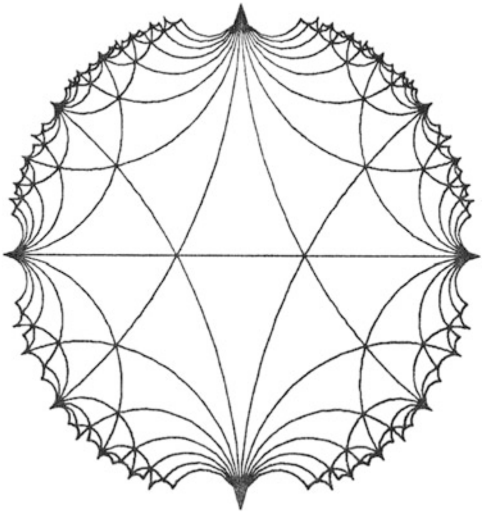


Fig. 3.20 Yet another tessellation of the unit disc. (Drawn by the USCD VAX computer and Mark Eggert)

imaginary quadratic number fields as we do in the next subsection. We will consider Minkowski's construction of a fundamental domain in $\mathcal{SP}_n$ for $GL(n,\mathbb{Z})$ in Volume II [667] as well as the analogue for more general discrete groups.

3.3.3 Computation of Class Numbers of Imaginary Quadratic Fields

Suppose that $K = \mathbb{Q}(\sqrt{d})$, $d < 0$, with $d =$ the discriminant of K. There is a one-to-one correspondence between ideal classes $C \in I_K$ (as defined in Sect. 1.4) and points z in the fundamental domain D for $SL(2,\mathbb{Z})$ in Fig. 3.14. For we can choose an ideal $\mathfrak{A}$ in C of the form

$$\mathfrak{A} = \mathbb{Z} \oplus \mathbb{Z}z, \quad z = (-b + \sqrt{d})/2a,$$

with

$$d = b^2 - 4ac, \quad a > 0, \quad \gcd(a,b,c) = 1.$$

Here a,b,c are in $\mathbb{Z}$. Thus for $z = (-b + \sqrt{d})/2a$ to be in the fundamental domain, we need $-a < b \le a$ and $1 \le |z|^2 = c/a$. It follows from the definition of the fundamental domain that the class number of K is

$$h_K = \#\left\{ (a,b,c) \in \mathbb{Z}^3 \;\middle|\; \begin{array}{c} \gcd(a,b,c) = 1,\ a > 0, \quad -a < b \le a,\ c \ge a, \\ c > a \text{ if } b < 0, \quad d = b^2 - 4ac \end{array} \right\}.$$

Table 3.2 Class numbers of imaginary quadratic fields

Discriminant	Class number	Discriminant	Class number
$-1,000,003$	105	$-1,000,052$	306
$-1,000,007$	630	$-1,000,055$	828
$-1,000,011$	368	$-1,000,056$	364
$-1,000,015$	430	$-1,000,059$	240
$-1,000,019$	342	$-1,000,063$	394
$-1,000,020$	320	$-1,000,067$	318
$-1,000,023$	706	$-1,000,068$	372
$-1,000,024$	274	$-1,000,072$	264
$-1,000,027$	168	$-1,000,079$	974
$-1,000,031$	928	$-1,000,083$	184
$-1,000,036$	192	$-1,000,084$	300
$-1,000,039$	877	$-1,000,087$	366
$-1,000,040$	688	$-1,000,088$	372
$-1,000,043$	192	$-1,000,091$	342
$-1,000,047$	508	$-1,000,095$	720
$-1,000,051$	276	$-1,000,099$	187

Note that you need only count the pairs (a,b), since they determine c. Also, the inequalities force $a < \sqrt{|d|/3}$. Thus the set of pairs to be considered is at most $|d|$ in number. It is an easy matter to program a computer to find the class number of an imaginary quadratic field K with d near -10^6. Table 3.2 gives some class numbers in this range. Shanks [586] used Gaussian composition of the quadratic forms $ax^2+2bxy+cy^2$ (which corresponds to multiplication of ideal classes) to speed the algorithm and found, for example, that if the discriminant is $-4,722,366,483,281,962,074,113$, then $h_K = 50,866,650,112$. See also Cohen [95] and Stein [637].

Exercise 3.3.5. Make a table of class numbers of imaginary quadratic fields similar to Table 3.2. I wrote the program to do the table on a programmable calculator in the early 1970s. How times have changed.

3.3.4 Dirichlet or Poincaré Polygon or Normal Fundamental Region

There is another construction of the fundamental region. It could be called the perpendicular bisector method since the boundaries are formed by perpendicular bisectors of geodesics joining a point w and γw, $\gamma \in \Gamma$, where w is assumed not to be fixed by any element of Γ. The resulting fundamental domain is often called the Dirichlet (or Poincaré) polygon or the normal fundamental region. The name comes because Dirichlet used the method for Euclidean space in 1850. Poincaré then used it for hyperbolic spaces. See Beardon [31], Svetlana Katok [342], and Maass [437].

Exercise 3.3.6 (Another Construction of a Fundamental Domain for $\Gamma\backslash$H). Suppose that Γ is a discrete subgroup of $SL(2,\mathbb{R})$ and Γ is infinite.

(a) Show that Γ is countable.
(b) Show that there is a point $w \in H$ which is not fixed by any element of Γ.
(c) Let $d(z_1,z_2)$ be the non-Euclidean distance between z_1 and z_2. Set

$$D = \{z \in H \ \mid \ d(z,w) \leq d(z,\gamma w) \text{ for all } \gamma \in \Gamma\},$$

where w is a point not fixed by any element of Γ. Show that D is a fundamental domain for Γ.

Hints.

(a) Any neighborhood of the identity in $SL(2,\mathbb{R})$ contains only finitely many points of Γ.
(c) Using the invariance of the non-Euclidean distance under elements of Γ, it is easy to show that H is a union of images of D under elements of Γ. Secondly, one must show that z and γz in D imply that z and γz lie on the boundary of D. Now the boundary of D consists of points z such that $d(z,w) = d(z,\lambda w)$ for some $\lambda \in \Gamma$, while the inequality holds for other $\gamma \in \Gamma$. In fact, if we set $D_\gamma = \{z \mid d(z,w) \leq d(z,\gamma w)\}$, then D_γ is a hyperbolic half-plane with a boundary that is the non-Euclidean perpendicular bisector of the geodesic through w and γw. Since D is the intersection of all the D_γ, $\gamma \in \Gamma$, it is a convex set in the sense of the non-Euclidean metric. If z and γz both lie in D, then $d(z,w) = d(z,\gamma^{-1}w)$ and z is a boundary point unless $\gamma = \pm I$. For more hints, see Maass [437, pp. 12–15], or Siegel [596, Vol. II, pp. 35–38]. It turns out that D may not have a finite number of sides. If it does, then it has finite area and conversely (see Siegel [596, Vol. II, pp. 39–46]).

We will call the fundamental region in Exercise 3.3.6 the **Dirichlet or Poincaré polygon** or **normal fundamental region**. In the special case $\Gamma = SL(2,\mathbb{Z})$, the normal fundamental region with center iy_0, $y_0 > 1$, is just the fundamental domain in Fig. 3.14. The non-Euclidean perpendicular bisectors of the lines joining iy_0 and $\pm 1 + iy_0$ are the vertical lines bounding this fundamental domain. The non-Euclidean perpendicular bisector of the geodesic joining iy_0 and i/y_0 is the circle of center 0 and radius 1, the other boundary of the fundamental domain. We mentioned an application of the computation of this fundamental domain in Sect. 3.3.3. There is also an application to the theory of Riemann surfaces of genus 1 or elliptic curves. For this fundamental domain is in correspondence with the classes of conformally equivalent Riemann surfaces of genus 1 (see Maass [437, pp. 51ff]).

Exercise 3.3.7. Construct fundamental domains for some congruence subgroups of $SL(2,\mathbb{Z})$. These subgroups are defined by formula (3.51), which follows.

The website of Helena Verrill has a wonderful fundamental domain drawing program

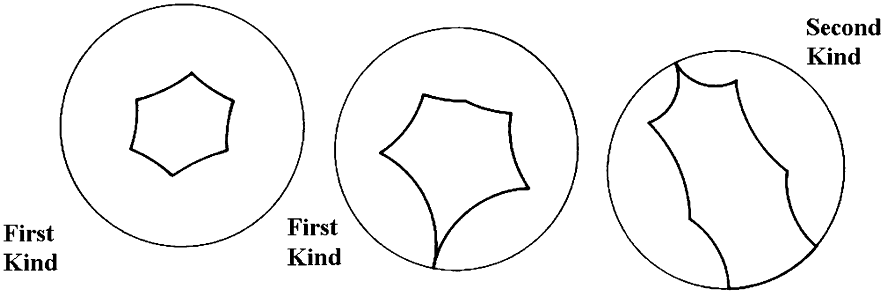


Fig. 3.21 Fundamental domains for Fuchsian groups. (Reprinted from Lehner, [410, pg. 22], by permission of the American Mathematical Society)

https://www.math.lsu.edu/~verrill/fundomain/magmaFD.html.

3.3.5 Other Discrete Subgroups of $SL(2,\mathbb{R})$

There is an overabundance of terminology for groups of fractional linear transformations mapping the inside of a fixed circle onto itself. We owe this perhaps to the rivalry of Poincaré and Klein (see their letters in *Acta Math. 39* (1923) or Rankin [533]). Poincaré called a group of fractional linear transformations mapping the inside of a fixed circle onto itself a **Fuchsian group**, and Klein called such a group a **Hauptkreisgruppe** (principal circle group).

There is a finer distinction, according to whether all points of the principal circle are limit points of the group. By a **limit point** of Γ, we mean a point of accumulation of the set of centers of the **isometric circles**

$$A_\gamma = \{z \in \mathbb{C} \mid |cz+d| = 1\}, \text{ for } \gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma \text{ and } c \neq 0.$$

If all points of the principal circle are limit points, the group is called a **Fuchsian group of the first kind** in English and a **Grenzkreisgruppe** in German. One can prove that Γ is a finitely generated Fuchsian group of the first kind if and only if Γ is a discontinuous subgroup of the automorphism group of the interior of its principal circle having a fundamental domain of finite Poincaré area. For our purposes the word "discontinuous" is equivalent to discrete. See Maass [437], for the definition of discontinuous group.

Exercise 3.3.8. Show that the modular group is a Fuchsian group of the first kind.

Poincaré called a subgroup of $SL(2,\mathbb{C})$ **Kleinian** if it had no limit circle. Klein apparently did not like this use of his name. Figure 3.21 shows some possible fundamental domains for Fuchsian groups.

We have already noted that the sides of our fundamental domain for $SL(2,\mathbb{Z})$ are identified by the generators of $SL(2,\mathbb{Z})$. This is also true in the general case of a fundamental domain with a finite number of sides (see Beardon [31], Svetlana Katok [342], and Maass [437, p. 26]). Poincaré reversed the process which creates the fundamental domain from a Fuchsian group and its generators. His theorem starts with a convex non-Euclidean polygon and its side-pairing maps and generates a Fuchsian group. See Beardon [31, Sect. 9.8].

Another important example of a discrete subgroup of $SL(2,\mathbb{R})$ is the **principal congruence subgroup** $\Gamma(N)$ of level N defined by

$$\Gamma(N) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \middle| \begin{array}{c} a,\ b,\ c,\ d \in \mathbb{Z}, \\ a \equiv d \equiv 1 (\operatorname{mod} N) \\ b \equiv c \equiv 0 (\operatorname{mod} N) \end{array} \right\}. \tag{3.51}$$

A reference on generators of this group for $N \equiv$ prime is Frasch [184]. A subgroup Γ of $SL(2,\mathbb{Z})$ is called a **congruence subgroup** if and only if Γ contains $\Gamma(N)$ for some positive integer N. It was discovered in the 1880s that there are an infinite number of examples of noncongruence subgroups (see Maass [437, pp. 76–78]). It turns out that $SL(n,\mathbb{Z})$ behaves quite differently for $n \geq 3$. For it has been proved by Mennicke as well as Bass, Lazard, and Serre that every subgroup of $SL(n,\mathbb{Z})$ of finite index is a congruence subgroup, if $n \geq 3$ (see Bass [29]). Thus the **congruence subgroup problem** was solved for $SL(n,\mathbb{Z})$.

Exercise 3.3.9. Show that a fundamental domain for a subgroup Γ_1 of $\Gamma = SL(2,\mathbb{Z})$ can be obtained by taking translates of a fundamental domain for Γ by representatives of the quotient Γ/Γ_1; i.e., if D is a fundamental domain for Γ, then a fundamental domain for Γ_1 is

$$\bigcup_{\gamma\Gamma_1 \in \Gamma/\Gamma_1} \gamma D.$$

Exercise 3.3.10. Show that if $\Gamma_1 = \Gamma(N)$, then it is a normal subgroup of $\Gamma = SL(2,\mathbb{Z})$ and the quotient Γ/Γ_1 is isomorphic to $SL(2,\mathbb{Z}/N\mathbb{Z})$. The only nontrivial part of this problem is the proof that for each solution $a,b,c,d \in \mathbb{Z}$ of $ad - bc \equiv 1$ $(\operatorname{mod} N)$, there are integers $a_1 \equiv a$, $b_1 \equiv b$, $c_1 \equiv c$, $d_1 \equiv d$ $(\operatorname{mod} N)$ such that $a_1 d_1 - b_1 c_1 = 1$.

It is also possible to show that the index of $\Gamma(N)$ in $SL(2,\mathbb{Z})$ is

$$[SL(2,\mathbb{Z}) : \Gamma(N)] = N^3 \prod_{p|N} (1 - p^{-2}),$$

where the product is over primes p dividing N. This result is proved in Shimura [589, pp. 21–22] as well as in Schoeneberg [562, pp. 74–75]. A fundamental domain for $\Gamma(2)$ is pictured in Fig. 3.22. Since

$$[SL(2,\mathbb{Z}) : \Gamma(2)] = 6,$$

Fig. 3.22 A fundamental domain for $\Gamma(2)$

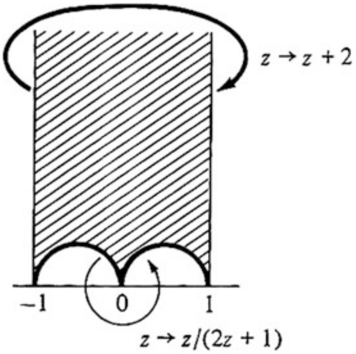
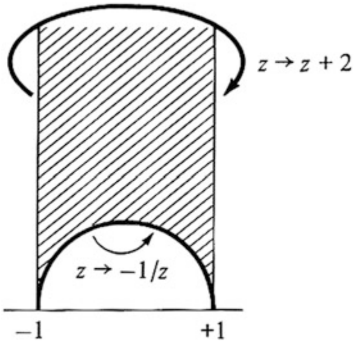


Fig. 3.23 Fundamental domain for the theta group



we can take our fundamental domain for $\Gamma(2)$ to be the union of six appropriate translates of a fundamental domain for $SL(2,\mathbb{Z})$. More information on fundamental domains for $\Gamma(N)$ can be found in Maass [437, pp. 71–73] and Schoeneberg [562, pp. 83–84]. See also Helena Verrill's website.

Recall from the formulas following Eq. (3.50) that the theta group is the group generated by $z \mapsto z+2$ and $z \mapsto -1/z$. A fundamental domain for the theta group is pictured in Fig. 3.23 (see Schoeneberg [562, pp. 84–86] and Maass [437, pp. 74–75]).

3.3.6 Riemann Surface of the Fundamental Domain

The final step in the study of fundamental domains for discrete subgroups of $SL(2,\mathbb{R})$ is to note that they can be made into **Riemann surfaces** with the upper half-plane as a (usually branched) covering surface.

First, we need to classify the fractional linear maps on the upper half-plane by classifying the corresponding elements of $SL(2,\mathbb{R})$ according to their Jordan form type (over $\mathbb{C}$). The Jordan form of such a matrix is one of the following:

$$\begin{pmatrix} a & 1 \\ 0 & a \end{pmatrix} \quad \text{or} \quad \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix}.$$

The corresponding fractional linear map will be one of four types.

The four types of fractional linear map

(1) $z \to z$ **identity**
(2) $z \to cz,\ |c| = 1,\ c \neq 1;$ **elliptic**
(3) $z \to cz,\ c > 0,\ c \neq 1;$ **hyperbolic**
(4) $z \to z + a, \quad a \neq 0.$ **parabolic**

In case (1), the element γ of $SL(2,\mathbb{R})$ is $\pm I$, $I =$ the identity. In case (2), the element γ is called elliptic. In case (3), the element γ is called hyperbolic. In case (4), the element γ is called parabolic.

Exercise 3.3.11. Show that if $\sigma \in SL(2,\mathbb{R}),\ \sigma \neq \pm I$,

(a) σ is parabolic $\Leftrightarrow |\mathrm{Tr}\ \sigma| = 2$.
(b) σ is elliptic $\Leftrightarrow |\mathrm{Tr}\ \sigma| < 2$.
(c) Is hyperbolic $\Leftrightarrow |\mathrm{Tr}\ \sigma| > 2$.

More details on the classification of elements of $SL(2,\mathbb{R})$ and $SL(2,\mathbb{C})$ can be found in Schoeneberg [562, pp. 2–4], and Shimura [589, pp. 5–10], for example.

A point $z \in H$ is called **elliptic** (with respect to Γ) if there is an elliptic element of Γ fixing z. A point s of $\mathbb{R} \cup \{\infty\}$ is called a **cusp** (with respect to Γ) if there is a parabolic element of Γ fixing s. These will be the interesting points on the Riemann surface of a fundamental domain for Γ. The elliptic points in the fundamental domain for $SL(2,\mathbb{Z})$ given in Fig. 3.14 are the points $i = \sqrt{-1}$, $\rho = (-1 + i\sqrt{3})/2$, and $\rho + 1 = -\overline{\rho}$. The last two points are really identified under the action of $SL(2,\mathbb{Z})$. The only cusp of this fundamental domain for $SL(2,\mathbb{Z})$ is the cusp at infinity.

For p in $H \cup \{\infty\}$, let S_p (the **stabilizer** of p) denote the group of maps $z \mapsto \gamma z$ which fix the point p, with $\gamma \in \Gamma$. For $\Gamma = SL(2,\mathbb{Z})$, $S_p =$ the identity, unless p is Γ− equivalent to ∞, i, or ρ. In these three cases one has the following: S_∞ is infinite cyclic generated by $z \mapsto z + 1$; S_i is cyclic of order 2 generated by $z \mapsto -1/z$; S_ρ is cyclic of order 3 generated by $z \mapsto -1/(z+1)$.

Exercise 3.3.12. Check the preceding statements about S_p for $\Gamma = SL(2,\mathbb{Z})$ and any $p \in H$.

The cusps and elliptic points often make trouble in calculations. We will experience this first-hand in our work on the Selberg trace formula. Thus some computations have been done for congruence subgroups $\Gamma(N)$ rather than for $SL(2,\mathbb{Z})$ itself, since these subgroups do not have elliptic points when $N > 1$ (see Exercise 3.3.13 and Fig. 3.22).

One can show that compactness of the fundamental domain $\Gamma \backslash H$, for Γ a discrete subgroup of $SL(2,\mathbb{R})$ implies that Γ has no parabolic elements. Examples of such groups having arithmetic interest come from quaternion algebras (see the discussion at the end of this section as well as Montserrat Alsina and Pilar Bayer [4], Fricke and Klein [185, pp. 502–634], Gelfand et al. [203, pp. 116–119], Svetlana Katok [342], Magnus [446], or Marie-France Vignéras [699]).

Exercise 3.3.13. Show that the congruence subgroups $\Gamma(N)$ of $SL(2,\mathbb{Z})$, with $N > 1$, have no elliptic elements.

Hint. *See Shimura [589, p. 22].*

For many purposes one needs the compactification of the fundamental domain, thought of as a Riemann surface. For the modular group, we need only add the point at infinity. This point is called a cusp. Of course, a Riemann surface needs local coordinate patches at every point. The only problem in describing the **local coordinates for the Riemann surface $D^* = D \cup \{\infty\}$, where D is the fundamental domain** of Fig. 3.14 **for $SL(2,\mathbb{Z})$**, occurs at the elliptic points and cusps. At any point $p \in D^*$ such that p is not an elliptic point or cusp, define the local coordinate t_p by $t_p = (z-p)/(z-\overline{p})$. When $p = i$, define the coordinate to be $t_i = [(z-i)/(z+i)]^2$. There is a double-branching here. When $p = \rho = (-1+i\sqrt{3})/2$ is the other elliptic point, define $t_\rho = [(z-\rho)/(z-\overline{\rho})]^3$. There is a triple-branching here. Then finally when $p = \infty$, define $t = \exp(2\pi i z)$. There is an infinite branching here. If you think about the fundamental domain and its identifications and try to draw neighborhoods of each point you will believe these "weird" coordinates are necessary and perhaps not even "weird." More details can be found in Maass [437, pp. 30–37], Schoeneberg [562, p. 27], and Shimura [589, p. 18].

One can now compute the genus of the branched Riemann surface D^* formed from the fundamental domain in Fig. 3.14 for $SL(2,\mathbb{Z})$. The genus is the number of handles when you view the Riemann surface as a sphere with handles. Now the fundamental domain is a non-Euclidean triangle. Thus you can use Euler's formula for the genus to see that D^* has **genus zero**. It is also possible to use a formula of Hurwitz to compute the genus of fundamental domains of other discrete subgroups of $SL(2,\mathbb{R})$ (see Svetlana Katok [342, p. 91], Lehner [410, Chap. 6], Maass [437, pp. 30–32], Schoeneberg [562, pp. 93–103], and Shimura [589, pp. 18–23]).

Exercise 3.3.14 (The Leaky Torus) (from Gutzwiller [239]). Consider the fundamental domain D in Fig. 3.24 with its edges identified by

$$A = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} \text{ and } B = \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix}.$$

(a) Show that the domain D, with sides identified as in Fig. 3.24, is equivalent to that given by Fig. 3.26, with boundary edges identified by A and B, plus

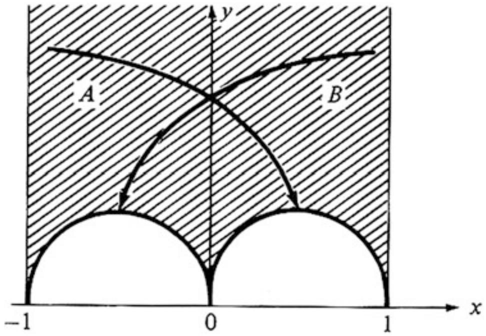


Fig. 3.24 Fundamental domain for the group in Exercise 3.3.14—the leaky torus. From Gutzwiller [239]. Reprinted by permission of Elsevier

$$K = \begin{pmatrix} -1 & -6 \\ 0 & -1 \end{pmatrix} = B^{-1}A^{-1}BA \quad \text{and} \quad C = \begin{pmatrix} 3 & -1 \\ 1 & 0 \end{pmatrix}.$$

Does it bother you that the group Γ generated by A and B is not the congruence group $\Gamma(2)$ although the pictures of the fundamental domains in Figs. 3.22 and 3.24 look the same? Can you explain how different discrete groups can have the same fundamental domain except for the boundary identifications?

(b) Show that the genus of the domain D in Fig. 3.24 is 1, while that in Fig. 3.22 is 0. Thus the fundamental domain for Γ generated by A and B above is topologically a torus with one cusp. This is pictured as the lower surface in Fig. 3.25. On the other hand, the fundamental domain for $\Gamma(2)$ is topologically a sphere with two cusps, which is pictured as the top surface in Fig. 3.25. The Riemann surface for the fundamental domain of the modular group is pictured as the center surface in Fig. 3.25.

(c) Does the group Γ generated by A and B above contain some congruence group $\Gamma(N)$? This would make Γ a congruence group of level N.

Hint. *See Schoeneberg [562] for hints on part (c).*

Gutzwiller [239, pp. 345–346], notes: “The leaky torus is topologically different from the ordinary torus which physicists like to use in Euclidean space. The four corners of the domain D become one point as usual, but this point is now infinitely removed. A path which goes around this exceptional point cannot be contracted to zero, because that would require moving it over an infinite distance.”

3.3.7 Triangle Groups and Quaternion Groups

There are discrete groups acting on H called **triangle groups**. A **Schwarz triangle group** (p,q,r) is generated by reflections in the sides of a geodesic triangle in H with angles $\pi/p, \pi/q, \pi/r$. Such groups make sense in the Euclidean plane as well as the sphere. In order that the group be hyperbolic, it is necessary that $\frac{1}{p}+\frac{1}{q}+\frac{1}{r}$ be

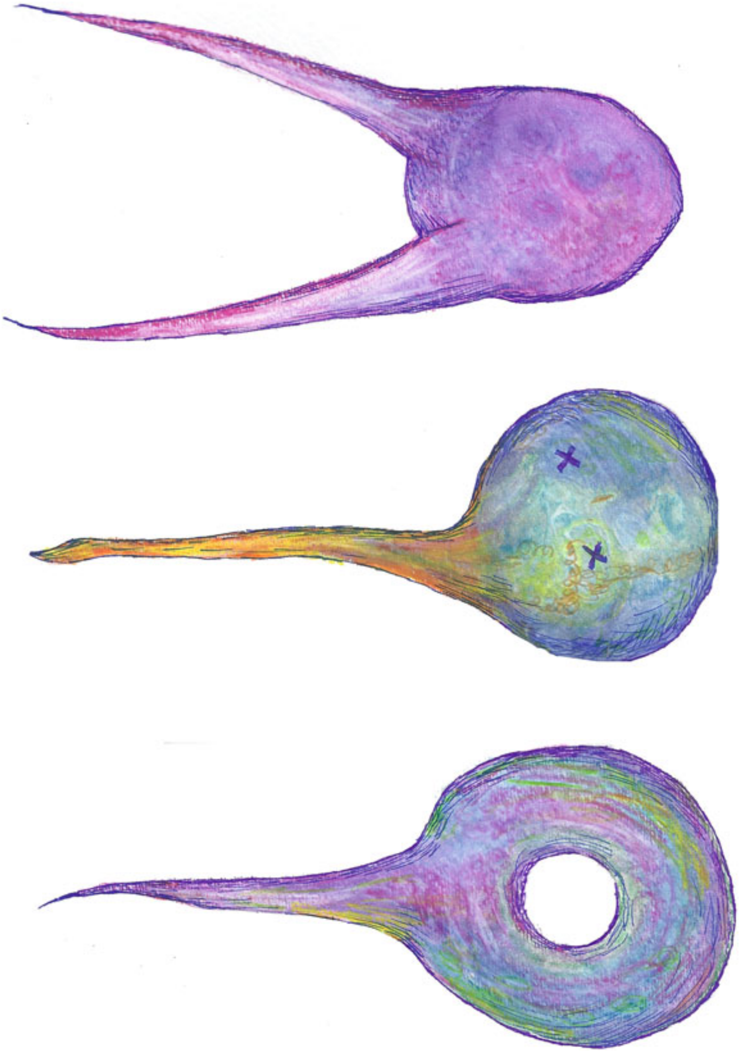


Fig. 3.25 Comparison of the surfaces for the leaky torus on the *bottom*, the fundamental domain of the modular group in the *center*, and the fundamental domain for $\Gamma(2)$ on the *top*. The x's in the fundamental domain in the center represent the branch points

less than 1. The modular group is the orientation-preserving elements of $(2,3,\infty)$. It is a subgroup of index 2 in the triangle group. The nontrivial coset is represented by the map $x+iy \to -x+iy$; i.e., the reflection in the vertical side. Figure 3.19 shows a tessellation of the unit disc for the modular group. The union of any shaded and white region gives a fundamental domain for the action of the modular group. Any individual triangle is a fundamental domain for the triangle group with $(2,3,\infty)$. A **Hecke triangle group** is similarly the orientation-preserving subgroup of $(2,q,\infty)$, with $q = 3,4,5,6,\ldots$.

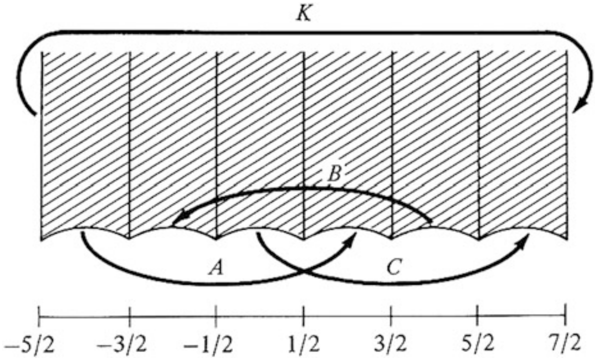


Fig. 3.26 Another fundamental domain for the leaky torus

Figure 3.18 shows a tessellation of the unit disc for the triangle group $(2,3,7)$. The group $(2,3,7)$ is the Klein quartic. The fundamental domain can be identified with the projective plane curve $x^3y + y^3z + z^3x = 0$. See Clark and Voight [91, p. 39], who construct subgroups of the triangle groups they call congruence subgroups.

A Fuchsian group of the first kind is called **arithmetic** if it is commensurable with a quaternion group coming from a quaternion algebra over a totally real algebraic number field of finite degree. Two subgroups A and B of a group such as $SL(2,\mathbb{R})$ are said to be **commensurable** when $A \cap B$ has finite index in both A and B. Takeuchi [652] has shown that there exist only finitely many arithmetic triangle groups up to $SL(2,\mathbb{R})$-conjugation. You can find Takeuchi's list of 85 arithmetic triangle groups in [653] as well as Bogomolny [46, p. 80]. If the numbers p and q are finite and $r = \infty$, the only arithmetic triangle groups have $(p,q,r) = (2,3,\infty), (2,4,\infty)$, and $(2,6,\infty)$.

Surprisingly, one can tell whether a group Γ is arithmetic by looking at the spectrum of the non-Euclidean Laplacian Δ on $\Gamma\backslash H$. For example, Hejhal [273] found no *even* Maass cusp forms for nonarithmetic Hecke triangle groups, while the arithmetic groups do have even Maass cusp forms. And computations of physicists such as C. Schmit, show that the nearest neighbor spacings of eigenvalues of Δ on $\Gamma\backslash H$ for non-arithmetic groups Γ is that of random symmetric real matrices (often called GOE, for Gaussian Orthogonal Ensemble) while that for arithmetic groups Γ is Poisson (e^{-x}). See Bogolmony [46, p. 80] and our discussion of random matrix theory and arithmetic quantum chaos in Volume II [667] (or [670, p. 343]).

In the rest of this subsection we consider discrete subgroups of $SL(2,\mathbb{R})$ coming from quaternion algebras. We are particularly interested in those groups with compact fundamental domains. References for this subject are Montserrat Alsina and Pilar Bayer [4], Gelfand et al. [203], Svetlana Katok [342], David Kohel and Helena Verrill [363], Marie-France Vignéras [702], and Voight [708].

The fundamental domains that we are considering here are also called **Shimura curves** (once they are made into a Riemann surface).[3] Shimura curves are associated with an order in a quaternion algebra A over the rationals $\mathbb{Q}$. A **quaternion algebra** is just a four-dimensional vector space over $\mathbb{Q}$ (or whatever field is your favorite) with a multiplication which is associative but not commutative. If you take the basis of A over $\mathbb{Q}$ to be $\{1,i,j,k\}$, then the multiplication in A is defined by requiring that $i^2=a$, $j^2=b$, $k=ij=-ji$, plus the usual rules for algebras; e.g., the distributive, associative laws. We write $A=\left(\frac{a,b}{\mathbb{Q}}\right)$.

Example 3.3.1. Think of the **Hamiltonian quaternions**. $\mathbb{H}=\mathbb{R}\oplus i\mathbb{R}\oplus j\mathbb{R}\oplus k\mathbb{R}$, where $i^2=j^2=k^2=-1$, $ij=k=-ji$. This is the quaternion algebra $\mathbb{H}=\left(\frac{-1,-1}{\mathbb{R}}\right)$. It can be shown to be a **division algebra** or **skew field**. That is, every non-0 element $q=q_1+q_2i+q_3j+q_4k$ of $\mathbb{H}$, with $q_1,q_2,q_3,q_4\in\mathbb{R}$, has an **inverse** for multiplication given by $\overline{q}/Nrd(q)$, where the **conjugate** of $q=\overline{q}=q_1-q_2i-q_3j-q_4k$ and the **reduced norm** of q is $Nrd(q)=q\cdot\overline{q}=q_1^2+q_2^2+q_3^2+q_4^2$. It is an **Exercise** to check that $\overline{q\cdot q'}=\overline{q'}\cdot\overline{q}$ and that $\overline{q}/Nrd(q)$ is the multiplicative inverse of a non-zero element q. It follows that the Hamiltonian quaternions have no zero divisors, which would be non-zero elements $q\in\mathbb{H}$ such that there are elements $w\in V$ with $qw=0$.

If the quaternion algebra A is isomorphic to $\mathbb{Q}^{2\times 2}$, A is called **split**. In this case, A is not a division algebra since it has 0-divisors like $\begin{pmatrix}0&1\\0&0\end{pmatrix}$.

Exercise 3.3.15. Show that the quaternion algebra $\left(\frac{1,1}{\mathbb{Q}}\right)$, with the notation defined above, is split.

For an element $q=q_1+q_2x+q_3y+q_4z$, with $q_1,q_2,q_3,q_4\in\mathbb{Q}$ and $q\in\left(\frac{a,b}{\mathbb{Q}}\right)$, where $z=xy$, $x^2=a$ and $y^2=b$, define the **conjugate** of q to be $\overline{q}=q_1-q_2x-q_3y-q_4z$, as before and then define the **reduced norm** $Nrd(q)=q\overline{q}$, as in the Example of Hamilton quaternions.

Exercise 3.3.16. Check that $\overline{q\cdot q'}=\overline{q'}\cdot\overline{q}$ and $Nrd(q)=q\cdot\overline{q}=q_1^2-q_2^2a-q_3^2b+q_4^2ab$.

We can define a linear mapping from $\left(\frac{a,b}{\mathbb{Q}}\right)$ into 2×2 matrices with entries in $\mathbb{Q}(\sqrt{a})$ by defining φ on the basis of the algebra and extending by linearity:

$$\varphi(1)=I,\ \ \varphi(x)=\begin{pmatrix}\sqrt{a}&0\\0&-\sqrt{a}\end{pmatrix},$$

$$\varphi(y)=\begin{pmatrix}0&1\\b&0\end{pmatrix},\ \ \varphi(z)=\begin{pmatrix}0&\sqrt{a}\\-b\sqrt{a}&0\end{pmatrix}.$$

---

[3] The word "curve" may seem odd. We are supposed to think it is one-dimensional over $\mathbb{C}$.

Exercise 3.3.17. (1) Show that, with the preceding definition of the $\mathbb{Q}$-linear map φ on the basis of the quaternion algebra $\left(\frac{a,b}{\mathbb{Q}}\right)$, if $q = q_1 + q_2 x + q_3 y + q_4 z$, with $q_1, q_2, q_3, q_4 \in \mathbb{Q}$, $q \in \left(\frac{a,b}{\mathbb{Q}}\right)$, we get

$$\varphi(q) = \begin{pmatrix} q_1 + q_2\sqrt{a} & q_3 + q_4\sqrt{a} \\ b\,(q_3 - q_4\sqrt{a}) & q_1 - q_2\sqrt{a} \end{pmatrix}.$$

(2) Show that φ defines an algebra isomorphism from $\left(\frac{a,b}{\mathbb{Q}}\right)$ onto a $\mathbb{Q}$-subalgebra of the algebra of all 2×2 matrices with entries in $\mathbb{Q}(\sqrt{a})$. This means that you need to check φ preserves multiplication by scalars, vector addition and algebra multiplication, in particular.

It is possible to do number theory in quaternion algebras A over $\mathbb{Q}$ (or some other number field) in a similar way to that in number fields except that the non-commutativity of multiplication throws a few monkey wrenches into the works. We have seen that there are still norms. One also has an analogue of the ring of integers. These are called maximal orders in A. An **order** $\mathfrak{O}$ in A is a subring of A containing 1 such that, as a $\mathbb{Z}$-module, $\mathfrak{O}$ is finitely generated and contains a set of 4 vectors that are linearly independent over $\mathbb{Q}$. A **maximal order** is just an order that is not a proper subset of some other order in A.

Fix a maximal order $\mathfrak{O}$ in A. A **unit** $u \in \mathfrak{O}$ in A means that u^{-1} is also in $\mathfrak{O}$. Note that we mean that u^{-1} is both a right and a left inverse. The **units of reduced norm 1=** $\boldsymbol{U} = \left\{u | u, u^{-1} \in \mathfrak{O}, Nrd(u) = 1\right\}$ is a subgroup of the unit group that can be identified (using φ) with a subgroup Γ of $SL(2,\mathbb{R})$.

If $A = \left(\frac{1,1}{\mathbb{Q}}\right)$, then $\varphi(A) = \mathbb{Q}^{2\times 2}$. So we can take our maximal order $\mathfrak{O}$ in A to be $\mathbb{Z}^{2\times 2}$. Why is it maximal? The Shimura curve in the case that $A = \left(\frac{1,1}{\mathbb{Q}}\right)$ is just the usual fundamental domain D of the modular group thinking of D as a Riemann surface with its sides identified, cusps added, and branch point neighborhoods understood as described previously.

When the algebra A is not split, the corresponding $\Gamma\backslash H$ is compact. The construction depends on the choice of $\mathfrak{O}$ and the embedding of the units into 2×2 matrices.

Example 3.3.2. Here we just concentrate on one example from David Kohel and Helena Verrill [363, pp. 213–215]. They note that this example is a popular one in the literature. Let the nonsplit quaternion algebra be $A = \left(\frac{2,-3}{\mathbb{Q}}\right)$ with generators x, y over $\mathbb{Q}$ and $x^2 = 2, y^2 = -3, z = xy = -yx$. Take $\mathfrak{O} = \mathbb{Z} \oplus \left(\frac{x+z}{2}\right)\mathbb{Z} \oplus \left(\frac{1+y}{2}\right)\mathbb{Z} \oplus z\mathbb{Z}$. Then we represent x, y, z by matrices as follows:

$$x \mapsto \begin{pmatrix} \sqrt{2} & 0 \\ 0 & -\sqrt{2} \end{pmatrix}, \quad y \longmapsto \begin{pmatrix} 0 & 1 \\ -3 & 0 \end{pmatrix}, \quad z \mapsto \begin{pmatrix} 0 & \sqrt{2} \\ 3\sqrt{2} & 0 \end{pmatrix}.$$

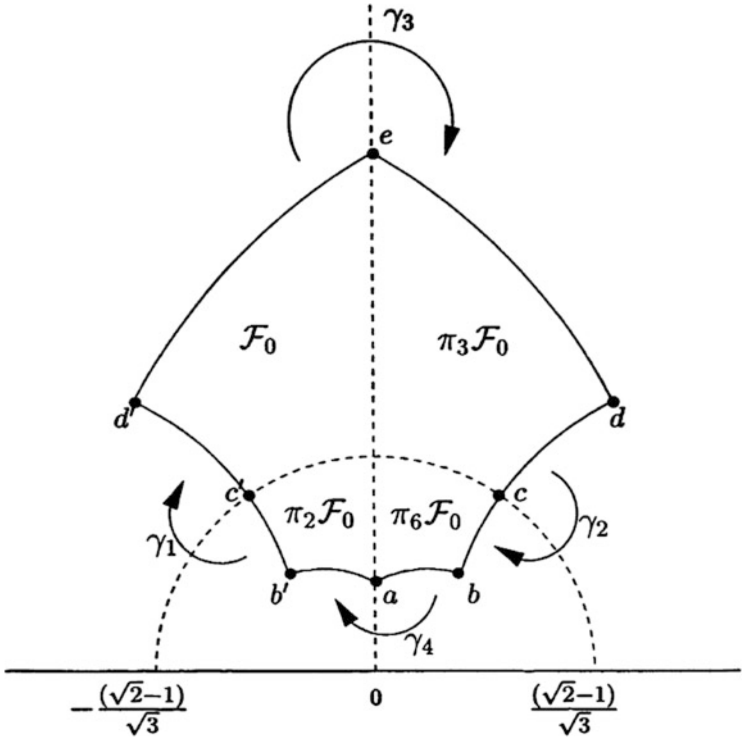


Fig. 3.27 Fundamental domain for an example of a quaternion group from David Kohel and Helena Verrill [363]

One must check that $\mathfrak{O}$ is a maximal order in A. Kohel and Verrill find a presentation of the group of units of positive norm in $\mathfrak{O}$: generators

$$\left\{ \begin{array}{c} \gamma_1 = (x+2y-z)/2,\ \gamma_2 = (x-2y+z)/2, \\ \gamma_3 = (1+y)/2,\ \gamma_4 = (1+3y-2z)/2 \end{array} \right\};$$

relations

$$\left\{ \gamma_1^2 = \gamma_2^2 = \gamma_3^2 = \gamma_4^2 = \gamma_4\gamma_2\gamma_3\gamma_1 = 1 \right\}.$$

They obtain the fundamental domain given in Fig. 3.27. The labelled points are

$$a = \frac{\left(-2\sqrt{2}+3\right)\sqrt{-3}}{3},\ \ b = \frac{\left(4\sqrt{2}-5\right)\left(3+2\sqrt{-3}\right)}{21},$$

$$c = \frac{\left(\sqrt{2}-1\right)\left(1+\sqrt{-2}\right)}{3}, \quad b' = -\overline{b}, c' = -\overline{c}.$$

The points d, d' and e are defined by $d' = \pi_6 b$, $d = \pi_6 b'$, and $e = \pi_6 a$, where $\pi_6 = 2y - z$, which is an element of the normalizer of Γ. Then $\varphi(\pi_6) = \begin{pmatrix} 0 & 2-\sqrt{2} \\ -6-3\sqrt{2} & 0 \end{pmatrix}$.

The Shimura curve corresponding to the preceding example is called $X_0^6(1)$. It is the fundamental domain of Fig. 3.27 with sides identified by the generators of the quaternion group in Example 3.3.2 and the neighborhoods of points given as we described earlier. One has a formula for the area of the fundamental domain F of a discrete group coming from a quaternion algebra of discriminant D given by

$$\text{area}(F) = \frac{\pi}{3}\prod_{p|D}(p-1),$$

where the product is over primes dividing D which is the discriminant of the maximal order in the quaternion algebra. This is proved by Marie-France Vignéras [702] . For the example of Fig. 3.27, the discriminant is 6^2. So we find the area is $2\pi/3$. You can check this numerically. The genus of this Riemann surface is 0. Ihara proved that this Shimura curve (i.e., the Riemann surface) can be represented as the solutions of the equation $X^2+Y^2+3Z^2 = 0$. See Kurihara [380].

The preceding results can be extended by replacing $\mathbb{Q}$ with any totally real algebraic number field. Shimura studied such groups since 1960 and viewed them, according to Montserrat Alsina and Pilar Bayer [4, p. xiii], as “moduli spaces of principally polarized abelian surfaces with quaternion multiplication and level structure.”

What sort of mathematical software should one use to do computations with discrete subgroups of $SL(2,\mathbb{R})$? Helena Verrill uses Magma to compute fundamental domains. See her chapter on subgroups of $SL(2,\mathbb{R})$ in *The Magma Handbook* [698]. Montserrat Alsina and Pilar Bayer use the Poincaré package in Maple to do their computations.

3.3.8 Finite Upper Half-Planes and Their Tessellations

We saw in Sect. 1.5 that it is useful to replace the real numbers $\mathbb{R}$ by a finite field $\mathbb{F}_q$ and the plane $\mathbb{R}^2$ with the finite plane $\mathbb{F}_q^2$. As in that section, we assume that q is odd throughout this discussion. So now we replace the Poincaré upper half-plane H with the **finite non-Euclidean “upper” half-plane** $H_q \subset \mathbb{F}_q(\sqrt{\delta})$ where δ is a nonsquare in the finite field $\mathbb{F}_q$ (for q odd) and

$$H_q = \left\{ z = x + y\sqrt{\delta} \,|\, x, y \in \mathbb{F}_q,\ y \neq 0 \right\}.$$

Here our finite analogue of the complex numbers is $\mathbb{F}_q(\sqrt{\delta}) = \mathbb{F}_{q^2}$. We have all the usual algebraic rules for computing in the field of complex numbers. For example, define, for $z = x + y\sqrt{\delta}$, the **imaginary part** of z to be $y = Im(z)$. Define the **conjugate** $\bar{z} = x - y\sqrt{\delta} = z^q$, and the **norm** $Nz = z\bar{z}$.

Note that H_q is can also be viewed as the union of an upper half-plane and a lower half-plane. We do not have a good analogue of positivity in a finite field. So we are just thinking that all non-0 elements are positive. You may prefer to think positive elements in $\mathbb{F}_q$ are the squares.

We define a **non-Euclidean "distance"** between $z, w \in H_q$ by

$$d(z,w) = \frac{N(z-w)}{Im(z)Im(w)}. \tag{3.52}$$

This is a natural finite analogue of the Poincaré metric for the classical hyperbolic upper half-plane. It is an exercise to show that $d(z,w)$ is invariant under fractional linear transformation

$$z \longrightarrow \frac{az+b}{cz+d}, \text{ with } \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in G = GL(2,\mathbb{F}_q) \text{ meaning that } ad - bc \neq 0.$$

Exercise 3.3.18. Let K be the subgroup of G fixing $\sqrt{\delta}$. Show that

$$K = \left\{ \begin{pmatrix} a & b\delta \\ b & a \end{pmatrix} \in G \right\} \cong \mathbb{F}_q(\sqrt{\delta})^*.$$

The subgroup K is a finite analogue of the group of real rotation matrices $O(2,\mathbb{R})$ in $GL(2,\mathbb{R})$.

Hint. *The mapping φ giving the isomorphism with the multiplicative group $\mathbb{F}_q(\sqrt{\delta})^* = \mathbb{F}_q(\sqrt{\delta}) - \{0\}$ is defined by setting*

$$\varphi \begin{pmatrix} a & b\delta \\ b & a \end{pmatrix} = a + b\sqrt{\delta}, \text{ for } a, b \in \mathbb{F}_q.$$

Exercise 3.3.19. Show that (G,K), with $G = GL(2,\mathbb{F}_q)$ and K as in the preceding exercise, gives a Gelfand pair and thus that $H_q \cong G/K$ is a finite symmetric space.

In the finite upper half-plane we can think of the set $\left\{ x + y\sqrt{\delta} \,\middle|\, y \in \mathbb{F}_q - 0 \right\}$, for fixed x in F_q, and its images under elements $g \in GL(2,\mathbb{F}_q)$ as **finite analogues of geodesics**. Similarly for fixed y in F_q, define $\left\{ x + y\sqrt{\delta} \,\middle|\, x \in \mathbb{F}_q \right\}$ and its images under elements $g \in GL(2,\mathbb{F}_q)$ to be **finite analogues of horocycles**.

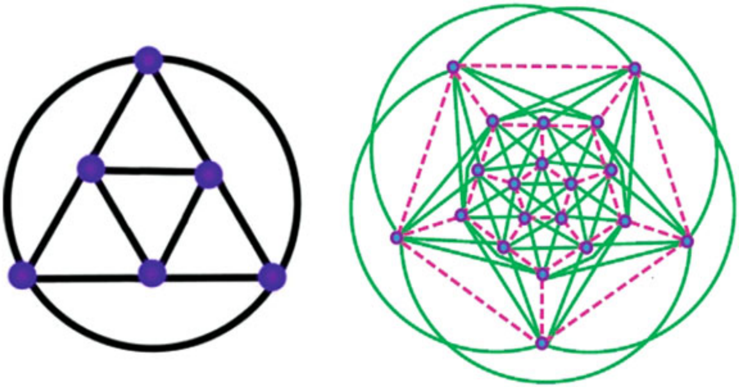


Fig. 3.28 Two finite upper half-plane graphs. On the *left* is $X_3(-1,1)$ and on the *right* is the graph $X_5(2,1)$ which has edges given by the *green lines*. The *magenta dashed lines* on the right are the edges of a dodecahedron on whose faces the finite upper half-plane graph gives *stars*. Of course, the large bottom face has a very distorted star

Fix an element $a \in \mathbb{F}_q$ with $a \neq 0, 4\delta$. Define the vertices of the **finite upper half-plane graph** $X_q(\delta, a)$ to be the elements of H_q. Connect vertex z to vertex w iff $d(z,w) = a$. See Fig. 3.28 for examples of such graphs.

Exercise 3.3.20.

(a) Show that the octahedron is the finite upper half-plane graph $X_3(-1,1)$.
(b) Draw the upper half-plane graph $X_9(\delta,1)$, where δ is a nonsquare in $\mathbb{F}_9 = \mathbb{F}_3\left(\sqrt{-1}\right)$.
(c) Show that $X_q(\delta, 4\delta)$ is a disconnected graph with edges connecting z and $\bar{z}$, for every $z \in H_q$.

Here we view finite upper half-planes as providing a “toy” symmetric space. The associated graphs give examples of Ramanujan graphs. An application to cryptography has also been found (see P.D. Tiu and Dorothy Wallace [681]).

The $GL(2,\mathbb{F}_q)$-invariant operators on H_q analogous to the non-Euclidean Laplacian on the Poincaré upper half-plane H are the **adjacency operators** on the finite upper half-plane graphs defined for $a \in \mathbb{F}_q$ by

$$A_a f(z) = \sum_{\substack{w \in H_q \\ d(z,w)=a}} f(w). \tag{3.53}$$

These operators generate a commutative algebra of operators.

Exercise 3.3.21. Define finite upper half-plane graphs over the finite fields of characteristic 2. The spectra of these graphs have been considered in Angel [9] and Evans [167].

It turns out (see Terras [668]) that simultaneous eigenfunctions for the adjacency operators of the finite upper half-plane graphs, for fixed δ as a varies over $\mathbb{F}_q$, are spherical functions for the symmetric space G/K, where K is the subgroup fixing $\sqrt{\delta}$.

Let $G = GL(2,\mathbb{F}_q)$. A **spherical function** $h : G \longrightarrow \mathbb{C}$ is defined to be a K bi-invariant eigenfunction of all the convolution operators by K-bi-invariant functions; it is normalized to have the value 1 at the identity. Here K-bi-invariant means $f(kxh) = f(x)$, for all $k,h \in K, x \in G$ and **convolution** means $f * h$, where

$$(f * h)(x) = \sum_{y \in G} f(y)h(y^{-1}x). \tag{3.54}$$

These spherical functions are analogues of Laplace (zonal) spherical harmonics (see Chap. 2) as well as the Legendre functions $h(ke^{-r}i) = P_{-\frac{1}{2}+it}(\cosh r)$ of Sect. 3.2, Theorem 3.2.2. Equivalently $h : H_q \to \mathbb{C}$ is a spherical function means h is a K-invariant eigenfunction for the adjacency matrices of all the graphs $X_q(\delta,a)$, as a varies over $\mathbb{F}_q$. One can show ([668], p. 347) that any **spherical function** has the form

$$h_\pi(x) = \frac{1}{|K|} \sum_{k \in K} \chi(kx), \tag{3.55}$$

where χ is the character of an irreducible unitary representation π of $G = GL(2,\mathbb{F}_q)$. Here π must appear in the induced representation (denoted $\rho = \mathrm{Ind}_K^G 1$) of the trivial representation of K induced up to G. This means $\rho = \mathrm{Ind}_K^G 1$ comes from the linear action of $g \in G$ on the space V of functions $V = \left\{f : G/K \cong H_q \to \mathbb{C}\right\}$ via $[\rho(g)f](z) = f(gz)$, for $z \in H_q$. In this situation (when G/K is a symmetric space or (G,K) a Gelfand pair) **eigenvalues = eigenfunctions**. See Stanton [621] and Terras [668, p. 344]. Thus our eigenvalues are also finite analogues of Legendre polynomials.

When π is a principal series representation π of G, we can write the spherical function h_π as a sum over K of the analogues of power functions $\lambda(\mathrm{Im}\, z)$, for λ a character of the multiplicative group $\mathbb{F}_q^*$. When π is a discrete series representation of G, the formula for the spherical function h_π is more complicated. We need both types of spherical functions to obtain an orthogonal basis for the space of all K-bi-invariant functions, as well as all the eigenvalues of the adjacency operators on the finite upper half-plane graphs. Soto-Andrade [611] managed to rewrite the sum (3.55), for the case π a discrete series representation of G, as an explicit exponential sum which is easy to compute (see formula (3.56) below). Thus we can call the discrete series eigenvalues of the adjacency matrices for the finite upper half-plane graphs **Soto-Andrade sums**. Katz [346] estimated these sums to show that the finite upper half-plane graphs are indeed Ramanujan graphs as defined in Sect. 1.5. Li [419] gives a different proof. Both proofs require the Riemann hypothesis for curves over finite fields, which was proved by Weil.

Suppose that ω is a nontrivial multiplicative character of the subgroup U of $\mathbb{F}_q\left(\sqrt{\delta}\right)^*$ consisting of elements t such that $Nt = t^{1+q} = 1$. Let ε be the character of the multiplicative group $\mathbb{F}_q^*$ that is 1 on squares, -1 on nonsquares, and 0 on 0. The **Soto–Andrade formula** for the spherical function associated to a discrete series representation of G if $d\left(z,\sqrt{\delta}\right) = a$ and $a \neq 0$ or 4δ, is given by:

$$(q+1)h_\omega(z) = \sum_{\substack{t=u+v\sqrt{\delta}\\ Nt=1}} \varepsilon\left(2(u-1)+\frac{a}{\delta}\right)\omega(t). \tag{3.56}$$

See [668, pp. 378 ff.] for a proof taken from Evans [168].

One can also view the eigenvalues of the finite upper half-plane graphs as entries in the character table of an association scheme. See Bannai [22]. And this whole subject can be reinterpreted in the language of Hecke algebras. See Krieg [373]. We will soon be discussing another sort of Hecke operator algebra.

Fundamental domains for subgroups Γ of $GL(2,\mathbb{F}_q)$ have been discussed by Shaheen [583] using the perpendicular bisector method. To do this, he needs to introduce an ordering of $\mathbb{F}_q$. Suppose that $g \in \mathbb{F}_q$ is a generator of the multiplicative group $\mathbb{F}_q^*$. **Order the elements of** $\mathbb{F}_q$ as follows:

$$0 < 1 < g < g^2 < g^3 < \cdots < g^{q-2}. \tag{3.57}$$

Of course this ordering will not have all the properties of the order in $\mathbb{R}$. And you could easily invent other orderings. For example, you can view elements of $\mathbb{F}_{p^2}$ as vectors $(x,y) \in \mathbb{F}_p \times \mathbb{F}_p$. You could use the order in formula (3.57) for elements of $\mathbb{F}_p$ and then the lexicographic order on the 2-vectors from $\mathbb{F}_p \times \mathbb{F}_p$.

Now Shaheen defines two **finite Dirichlet polygons**

$$D_\Gamma(z_0) = \left\{ w \in H_q \middle| d(z_0,w) < d(\gamma z_0,w), \forall \gamma \in \Gamma, \gamma \neq cI, \text{ for some } c \in \mathbb{F}_q \right\},$$

and

$$\overline{D_\Gamma(z_0)} = \left\{ w \in H_q \middle| d(z_0,w) \leq d(\gamma z_0,w), \forall \gamma \in \Gamma, \gamma \neq cI, \text{ for some } c \in \mathbb{F}_q \right\}.$$

Shaheen proves that for any $z \in H_q$, there exists $\gamma \in \Gamma$ such that $\gamma z \in \overline{D_\Gamma(z_0)}$. He also shows that if $z,w \in D_\Gamma(z_0)$ and $\gamma z = w$ for some $\gamma \in \Gamma$, it follows that $\gamma = cI$, for some $c \in \mathbb{F}_q$. This means that $D_\Gamma(z_0)$ has properties analogous to the Dirichlet polygon or normal fundamental region of Exercise 3.3.6.

Exercise 3.3.22. Do the finite analogue of part (c) of Exercise 3.3.6.

One can also consider $\Gamma-$tessellations of the finite upper half-plane H_q, where we assume $q = p^r$, with $r > 1$, and $\Gamma = GL(2,\mathbb{F}_p)$. We used the finite fields package in Mathematica to produce Fig. 3.29, which represents a tessellation of a fundamental domain for $GL(2,\mathbb{F}_{11})$ on H_{121}. A fundamental domain for $GL(2,\mathbb{F}_{11})$ can be taken to be any set of 11 colored squares.

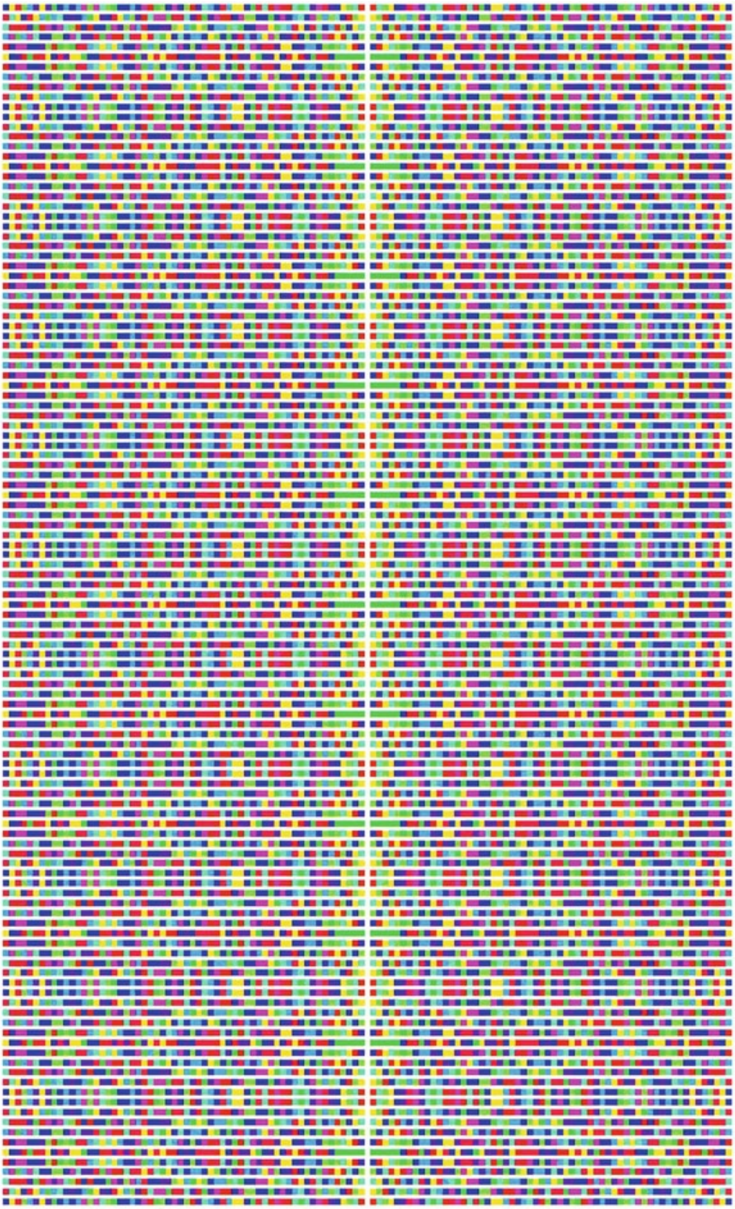


Fig. 3.29 Tessellation for $GL(2,\mathbb{F}_{11})$ acting on the finite upper half-plane over the field with $11 * 11$ elements. A fundamental domain consists of points with 11 distinct colors. The figure is rotated 90° and the *white line* down the middle denotes the analogue of the real axis, which is excluded from H_{121}

Exercise 3.3.23.

(a) Compute the fundamental domain for $GL(2,\mathbb{F}_5)$ acting on H_{25} and draw a tessellation of H_{25} given by this fundamental domain.

(b) Use the orbit-counting lemma (often called the Burnside lemma) to show that $GL(2,\mathbb{F}_p)\backslash H_{p^2}$ has p elements. This lemma says that if $Fix(\gamma) = \left\{ z\varepsilon H_{p^2} \middle| \gamma z = z \right\}$, then

$$\left|\Gamma\backslash H_{p^2}\right| = \frac{1}{|\Gamma|}\sum_{\gamma\varepsilon\Gamma} |Fix(\gamma)|.$$

3.4 Modular or Automorphic Forms-Classical

> The abundance of topics in this subject is so large that I have investigated only a few problems. I think that these facts could be put in order only by using notions from different arithmetical theories which are perhaps not sufficiently developed.
>
> —From Hecke [259].

3.4.1 *Definitions, Eisenstein Series*

There are many books on classical (by this I mean holomorphic or meromorphic) modular or, more generally, automorphic forms. To the list at the beginning of Sect. 3.3, you can add Diamond and Shurman [133], Iwaniec [319], Kilford [350], Koblitz [359], Ogg [505, 506], Sarnak [555], Stein [638], and Zagier [749], for example. Thus I do not feel compelled to write a book on this subject and we shall only present *some* of the classical theory in this section (and *none* of the modern adelic theory). Moreover, we will usually consider only the discrete group $SL(2,\mathbb{Z})$. For a treatment from the point of view of adelic representation theory, see Bernstein and Gelbart [37], Bump [71], Gelbart [200], Goldfeld and Hundley [219], or Jacquet and Langlands [323].

There are also many websites. One such is http://www.lmfdb.org/ for the LMFDB, which stands for the database of L-functions, modular forms, and related objects. Another is http://modform.org. William Stein's website has many useful things including books and computations with SAGE. Here we will use Mathematica mostly.

Definition 3.4.1. A function $f : H \to \mathbb{C}$ is said to be an **(entire) modular or automorphic form of weight *k* for the discrete group** $\Gamma = SL(2,\mathbb{Z})$ if

(1) f is entire on H; i.e., $f(z)$ is holomorphic for every $z \in H$;

(2) $f(\gamma z) = (cz+d)^k f(z)$, for any $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ in Γ;

(3) f is holomorphic at infinity; i.e., f has the Fourier series expansion

$$f(z) = \sum_{n \geq 0} a_n \exp(2\pi i n z).$$

Usually the word "modular form" refers to a function with an invariance property for $SL(2,\mathbb{Z})$ or one of its subgroups while we call functions with invariance properties for more general discrete subgroups of $SL(2,\mathbb{R})$ "automorphic forms." In order to check (2), it suffices to check that $f(z+1) = f(z)$ and $f(-1/z) = z^k f(z)$. Property (3) is equivalent to the requirement that f be bounded in the fundamental domain for $SL(2,\mathbb{Z})$.

The **vector space of holomorphic modular forms of weight k for $SL(2,\mathbb{Z})$** will be denoted $\boldsymbol{\mathcal{M}(SL(2,\mathbb{Z}),k)}$. *In order that this space be nonzero, the weight k must be a nonnegative even integer*. To see that k must be even, replace γ by $-\gamma$ in property (2). To see that k must be nonnegative, note that the function $y^{k/2}|f(z)|$ is invariant under $SL(2,\mathbb{Z})$. If $k < 0$, $y^{k/2}$ and thus $y^{k/2}|f(z)|$ is bounded on the fundamental domain and thus on all of the upper half-plane. Thus $|f(z)| \leq My^{-k/2}$ for some constant M. But then the coefficients in the Fourier series for f satisfy

$$a_n \exp(-2\pi n y) = \int_0^1 f(x+iy) \exp(-2\pi i n x) dx.$$

Thus $|a_n| \leq My^{-k/2} \exp(2\pi n y)$. If k is negative, it follows that $a_n = 0$ for all n, since you can let y approach zero (from above).

In order to obtain nonconstant modular forms of weight $k = 0$, one must relax the definition and allow the function f to have poles. Automorphic forms of weight zero are called **automorphic functions**. For groups Γ with more than one cusp, one must consider Fourier expansions at each cusp in part (3) of the definition of automorphic form. And in order to call the theta function in formula (3.50) of Sect. 3.3 a modular or automorphic form of weight $1/2$, one must introduce the concept of multiplier system in part (2) of the definition of modular form. A **multiplier system** $v : \Gamma \to \mathbb{C}$, corresponding to the weight k, has the properties $|v(\gamma)| = 1$ for all $\gamma \in \Gamma$, and

$$v(\gamma_1\gamma_2)(c_3 z + d_3)^k = v(\gamma_1)v(\gamma_2)(c_1\gamma_2 z + d_1)^k (c_2 z + d_2)^k \tag{3.58}$$

for

$$\gamma_i = \begin{pmatrix} a_i & b_i \\ c_i & d_i \end{pmatrix}, \; i = 1,2,3 \text{ and } \gamma_3 = \gamma_1\gamma_2.$$

If k is an integer, then Eq. (3.58) reduces to $v(\gamma_1\gamma_2) = v(\gamma_1)v(\gamma_2)$.

We say that $f : H \to \mathbb{C}$ is an **automorphic form of weight k and multiplier system v for Γ** if f satisfies the conditions (1) and (3) of the preceding definition (perhaps modified somewhat as we mentioned above), and condition (2) is replaced by

$$(2') \qquad f(\gamma z) = v(\gamma)(cz+d)^k f(z) \text{ for } \gamma = \begin{pmatrix} * & * \\ c & d \end{pmatrix} \in \Gamma.$$

Here, when the weight $k \notin \mathbb{Z}$, one must fix the branch of $(cz+d)^k$ by assuming that

$$z^k = |z|^k e^{ik\,\arg z}, \; -\pi \leq \arg z < \pi, \text{ for } z \in \mathbb{C}.$$

There is an overabundance of terminology with respect to the word "weight." Some replace k by $-k$ (cf. Lehner [410, p. 268]). Some use the word "degree" (cf. Knopp [357, p. 12]). See also Serre [576] for slightly divergent notation, as well as a nice short survey of the subject of this section—a survey which we follow somewhat here.

As an example of a modular form of weight $k = 4, 6, \ldots$, consider the **Eisenstein series G_k** defined by

$$G_k(z) = \sum_{(m,n)\in\mathbb{Z}^2-0} (mz+n)^{-k}, \; k = 4, 6, 8, \; \ldots. \tag{3.59}$$

Exercise 3.4.1.

(a) Set

$$\Gamma_\infty = \left\{ \begin{pmatrix} a & b \\ 0 & c \end{pmatrix} \in SL(2,\mathbb{Z}) \right\}.$$

Show that

$$G_k(z) = 2\zeta(k) \sum_{\gamma\in\Gamma_\infty\backslash SL(2,\mathbb{Z})} \left(\frac{d(\gamma z)}{dz} \right)^{k/2},$$

where $\zeta(k)$ is Riemann's zeta function. The sum is over a complete set of representatives $\gamma \in SL(2,\mathbb{Z})$ for the equivalence relation

$$\gamma_1 \sim \gamma_2 \text{ iff } \gamma_1 = \gamma\gamma_2, \text{ for some } \gamma \in \Gamma_\infty.$$

(b) Show that $G_k(z)$ is a nonzero modular form for $SL(2,\mathbb{Z})$ of weight k, if $k = 4, 6, 8, \ldots$.

Hints. *For the convergence of Eq. (3.59), compare the series with the Epstein zeta function in Exercise 1.4.5 of Sect. 1.4. To see that G_k has property (2) of modular* *forms, use part (a) and the chain rule. To see that* $G_k \neq 0$ *when* $k = 4, 6, 8, \ldots$*, show that*

$$\lim_{y \to \infty} G_k(iy) = 2\zeta(k),\ k = 4, 6, 8, \ldots.$$

Exercise 3.4.2. Obtain the complete Fourier expansion of the Eisenstein series G_k, $k = 4, 6, 8, 10, \ldots$.

Answer.

$$G_k(z) = 2\zeta(k) + 2\frac{(2\pi i)^k}{(k-1)!} \sum_{n \geq 1} \sigma_{k-1}(n) \exp(2\pi i n z),$$

where the ***divisor function*** *is*

$$\sigma_k(n) = \sum_{0 < d | n} d^k.$$

Here the sum is over the positive divisors of n.

Hint. *One method is to use the Poisson sum formula (see Maass [437, pp. 209–212] for a generalization). Another method starts with the partial fraction expansion of* $\pi \cot(\pi z)$ *(see Shimura [589, pp. 32–33]).*

Exercise 3.4.3 (The Discriminant Δ). Define[4] the **discriminant**

$$\Delta(z) = (60G_4)^3 - 27\,(140G_6)^2.$$

Show that Δ is a modular form of weight 12. Then prove that $\Delta(\infty) = 0$. Feel free to use Exercise 3.5.7 from Sect. 3.5. Graph a real-valued function related to $\Delta(z)$ with Mathematica or your favorite program. Compare with Fig. 3.30 where a Mathematica density plot of $(2\pi)^{-12} y^6 |\Delta(z)|$ is shown, using the relation $\Delta = (2\pi)^{12}\eta^{24}$, involving the Dedekind eta function defined below. See formulas (3.68) and (3.69) below. A website with many density plots of modular forms is that of Frank Farris. See the following page for density plots of various sums over the modular group produced by Frank Farris and Jeff Hoffstein:

http://math.scu.edu/~ffarris/auto/auto.html

The website

http://math.scu.edu/~ffarris/EdgeofUniverse.pdf

---

[4] Hopefully it will not be too confusing to use Δ both as the Laplace operator and as this modular form.

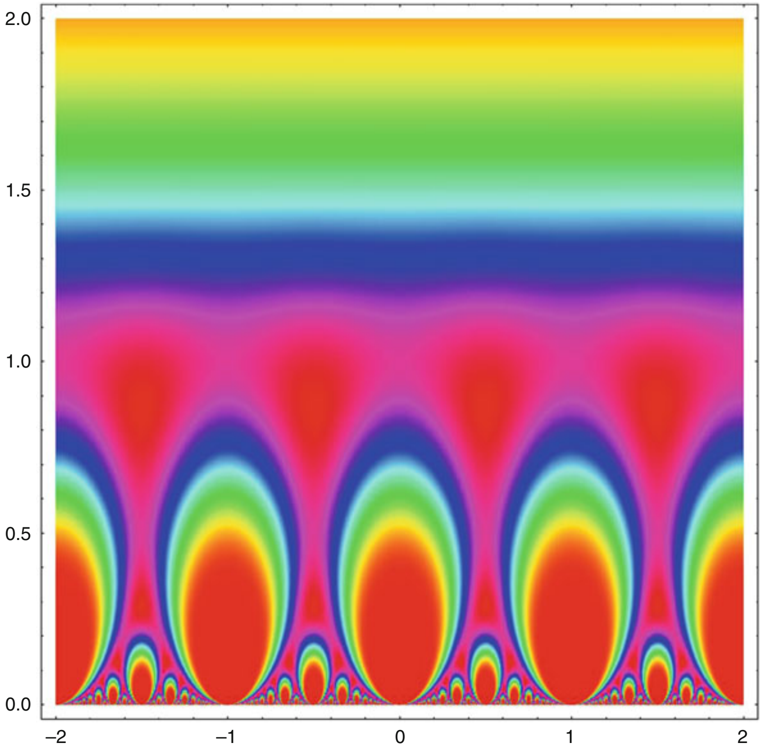


Fig. 3.30 This density plot of $(2\pi)^{-12}y^6\,|\Delta(z)|$ was produced with the Mathematica command: `DensityPlot[Abs[(y^6)*DedekindEta[x + I*y]^24], {x, -2, 2}, {y, 0, 2},ColorFunction -> Hue, PlotPoints -> 500]`

contains a color version of the Farris article [174] which did not appear in color on the M.A.A. website for the journal.

A modular form that vanishes at infinity is called a **cusp form** (*forme parabolique* in French or *Spitzenform* in German). We define **the space of (holomorphic) cusp forms of weight k,** $\boldsymbol{\mathcal{S}(SL(2,\mathbb{Z}),k)}$, to be the modular forms of weight k that vanish at infinity. From Exercise 3.4.3, we see that $\Delta \in \mathcal{S}(SL(2,\mathbb{Z}),12)$. However, it is possible that Δ is identically zero. There are many ways to see that this cannot happen. Look at a table of Fourier coefficients of Δ such as Table 3.3, for example. After we have more information on the orders of zeros of modular forms we will learn that Δ does not vanish at any point in H. See the proof of Theorem 3.4.1. This means that you can divide by Δ and that is a useful fact for class field theory (see Borel and Chowla [53]).

The facts about zeros of modular forms can be derived in several ways. First, if p is a point in the usual fundamental domain D for $SL(2,\mathbb{Z})$ as in Fig. 3.14, then let $n = n_p(f)$ denote the **order of the zero** of f at p; that is the integer n such that

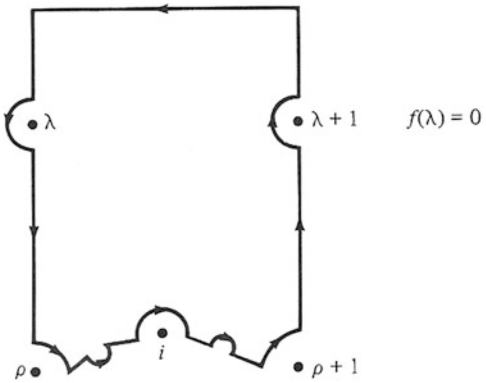


Fig. 3.31 The contour C needed for the proof of formula (3.60)

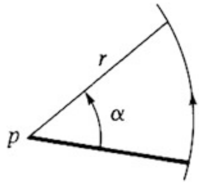


Fig. 3.32 The arc A near a zero of $f(z)$

$f(z)(z-p)^{-n}$ is holomorphic and nonzero at p. If f is in $\mathcal{S}(SL(2,\mathbb{Z}),k)$, then $n_p(f)=n_{\gamma p}(f)$ for all γ in $SL(2,\mathbb{Z})$. The **order of the zero of f at infinity** is denoted $n=n_\infty(f)$ and is obtained by looking at the Fourier expansion; i.e.,

$$f(z)=\sum_{r\geq n} a_r \exp(2\pi irz),\ a_n\neq 0.$$

Let e_z denote the order of the group of maps $z\mapsto \gamma z$ (with $\gamma\in\Gamma$) fixing $z\neq\infty$. Then $e_z=1$ when z is not $\Gamma-$equivalent to i or $\rho=e^{2\pi i/3}$; $e_i=2$ and $e_\rho=3$. The **fundamental formula for the orders of zeros of modular forms of weight k for $SL(2,\mathbb{Z})$** *is*

$$n_\infty(f)+\sum_{p\in D}\frac{n_p(f)}{e_p}=\frac{k}{12}, \tag{3.60}$$

where D is the standard fundamental domain for $SL(2,\mathbb{Z})$ (see Fig. 3.14).

The easiest proof of this formula comes from the Riemann–Roch theorem (see Gunning [236, Chap. II], and Shimura [589]). This method also generalizes to higher dimensions (see Yamazaki [747]). However, in order to keep this section as elementary as possible, we will present a proof by contour integration.

The proof will be carried out by integrating $d\log f(z)$ over the contour C in Fig. 3.31, taking the contour to enclose all zeros of f in the fundamental domain minus the points $i,\ \rho,\ \rho+1,\ \infty$. Note that f has only a finite number of zeros in the standard fundamental domain because f is holomorphic at infinity, so zeros cannot accumulate there.

To do the contour integral over C, first note that if $f(z) = (z-p)^m g(z)$ with $g(p) \neq 0$, then, when you integrate $d\log f(z)$ over the arc A in Fig. 3.32 and let the radius r go to zero, you get $m\alpha i$. There is cancellation of the curved parts of the contour integral of $d\log f(z)$ around the zeros of f on the vertical boundaries. And the curved part of the contour integral of $d\log f(z)$ around i contributes $-n_i\pi i$. Finally, the curved part of the contour integral around ρ contributes $-n_\rho\pi i/3$, with the same contribution for the curved part of the contour integral around $\rho+1$.

The limit of the part of the contour integral of $d\log f(z)$ along the horizontal contour at the top of C, as the horizontal line moves up to infinity, is $-2\pi i n_\infty$, using the expansion of $f(z)$ at infinity.

The contour integral over the bottom of the contour C is obtained from the fact that f is a modular form of weight k, as follows:

$$\begin{aligned}\int_{\text{bottom of contour}} d\log f(z) &= \int_i^\rho d\log f(z) - d\log f(-1/z)\\ &= \int_i^\rho (-k)\frac{dz}{z} = k\left(\frac{\pi}{2}-\frac{\pi}{3}\right) = \frac{k}{6}\pi i.\end{aligned}$$

Now, since the integrals over the vertical parts of the contour C cancel, one obtains the following formula for the number n of zeros of $f(z)$ in the fundamental domain minus the points $i,\ \rho,\ \rho+1,\ \infty$:

$$n = \frac{1}{2\pi i}\int_C d\log f(z) = -\frac{n_i}{2} - \frac{n_\rho}{6} - \frac{n_\rho}{6} - n_\infty + \frac{k}{12}.$$

Here the zeros are counted with multiplicity. This completes the proof of formula (3.60).

Formula (3.60) leads to the following theorem.

Theorem 3.4.1 (Formula for the Dimension of the Space of Holomorphic Modular Forms for SL(2, $\mathbb{Z}$)). *If $\lfloor x \rfloor$ denotes the floor of x; i.e., the greatest integer less than or equal to x, we have the following formula for the dimension of the space of modular forms of weight k for $SL(2,\mathbb{Z})$*

$$\begin{aligned}d_k &= \dim_{\mathbb{C}} \mathcal{M}(SL(2,\mathbb{Z}),k)\\ &= \begin{cases} 0, & \text{if } k<0 \text{ or } k\notin 2\mathbb{Z},\\ 1, & \text{if } k=0,\\ \left\lfloor \frac{k}{12}\right\rfloor +1, & \text{if } k\equiv 0,4,6,8,10 (\operatorname{mod} 12), \text{ for } k>0,\\ \left\lfloor \frac{k}{12}\right\rfloor, & \text{if } k\equiv 2 (\operatorname{mod} 12), \text{ for } k>0\end{cases}\end{aligned}$$

Proof. We have already proved the first line of the dimension formula (see the discussion immediately following the definition of $\mathcal{M}(SL(2,\mathbb{Z}),k)$).

Using formula (3.60), we find that $d_2 = 0$, since

$$\frac{1}{6} = n + n_\infty + \frac{n_i}{2} + \frac{n_\rho}{3}$$

is not solvable in non-negative integers n, n_∞, n_i, n_ρ.

It also follows that

$$d_k \leq 1 + \lfloor k/12 \rfloor . \tag{3.61}$$

For suppose that $m > 1 + \lfloor k/12 \rfloor$ and $f_1, \ldots, f_m \in \mathcal{M}(SL(2,\mathbb{Z}),k)$. Then there are constants $c_1, \ldots, c_m$, not all of which are zero, such that

$$\sum_{i=1}^{m} c_i f_i^{(j)}(\infty) = 0 = \quad \text{for } \ j = 0, 1, \ldots, m-2.$$

Thus $f = \Sigma_{i=1}^{m} c_i f_i$ is an element of $\mathcal{M}(SL(2,\mathbb{Z}),k)$ with a zero of order $\geq m-1$ at ∞. But then the formula for the orders of zeros of elements of $\mathcal{M}(SL(2,\mathbb{Z}),k)$ forces a contradiction. The second line of the dimension formula follows from Eq. (3.61).

We know from Exercises 3.4.1 and 3.4.2 that the Eisenstein series G_k, $k = 4, 6, 8, \ldots$, are nonzero elements of $\mathcal{M}(SL(2,\mathbb{Z}),k)$. This implies that $d_k = 1$ for $k = 4, 6, 8, 10$. Then Exercise 3.4.3 implies that $d_{12} = 2$, using formula (3.60) again. For we find from formula (3.60) that $n_\infty(\Delta) = 1$ and $n = n_i = n_p(\Delta) = 0$. Thus Δ *does not vanish at any finite point of* H. Note here that Δ is not identically 0, since $G_4(\rho) = 0 = G_6(i)$.

To finish the proof of the dimension formula, note that every form in $\mathcal{M}(SL(2,\mathbb{Z}),k)$ is a unique linear combination of forms $G_{k-12n}\Delta^n$, for $n \geq 0$, $k - 12n \geq 4$. To prove this, use the fact that there is a constant a_0 with $(f - a_0 G_k)(\infty) = 0$. Since Δ is only zero at ∞, you obtain $(f - a_0 G_k)/\Delta \in \mathcal{M}(SL(2,\mathbb{Z}),k-12)$. The proof is completed by induction. □

It follows from the preceding argument that every modular form of weight k is a unique linear combination of terms of the form $G_4^a G_6^b$, with nonnegative integers a, b such that $4a + 6b = k$. The proof is in Serre [576, p. 89].

More general results on dimensions of spaces of automorphic forms with multiplier systems for $SL(2,\mathbb{Z})$ can be found in Maass [437, p. 144]. Both Eisenstein series and Poincaré series (to be considered later in this section) are used to generate the spaces of such forms. The same result can also be found in Rankin [532, p. 208]. Both of these references show that there are exactly six different multiplier systems for $SL(2,\mathbb{Z})$ for each weight k (see Maass [437, p. 132] and Rankin [532, p. 206]). Rankin obtains the basis elements of the form $G_{r+12s}\Delta^{-s+(k-r)/12}, 0 \leq s \leq (k-r)/12$, with $r = 0, 4, 6, 8, 10, 14$ if $k \geq r$. Again, it is important that Δ does not vanish at any point (not ∞) in the upper half-plane. If $k < r$, then there are no forms but 0. If $k \not\equiv r (\text{mod } 12)$, the forms above are all cusp forms. A dimension formula for automorphic forms for more general discrete groups is to be found in Shimura [589, pp. 46–47]. The computation is a nice application of the Riemann–Roch theorem for curves.

Next we want to consider some examples of modular forms—the discriminant Δ, the modular invariant J, theta functions, etc.

3.4.2 The Discriminant, the Weierstrass $\wp$-Function, and Ramanujan's Tau Function

The **discriminant** $\Delta(z)$ is defined by the formula in Exercise 3.4.3. It will be convenient to use the notation

$$g_2 = 60 G_4 \quad \text{and } g_3 = 140 G_6. \tag{3.62}$$

The function Δ is called the discriminant because it is the discriminant of the cubic on the right-hand side of the equation

$$u'^2 = 4u^3 - g_2 u - g_3, \tag{3.63}$$

satisfied by the Weierstrass elliptic function $u = \wp(t)$ (see Ahlfors [3, pp. 264–268] or Koblitz [359]). The cubic equation in formula (3.63) is an elliptic curve. We will say more about these curves later.

Define the **Weierstrass function** $\wp(t) = \wp(t,z)$ for $t \in \mathbb{C}$ and $z \in H$ by

$$\wp(t,z) = \frac{1}{t^2} + \sum_{(m,n)\in\mathbb{Z}^2-0} \left(\frac{1}{(t-m-nz)^2} - \frac{1}{(m+nz)^2} \right). \tag{3.64}$$

The function $\wp(t,z)$ is really associated to the period lattice $L = \mathbb{Z} \oplus z\mathbb{Z}$. Sometimes $\wp$ is written $\wp(t,z) = \wp(t;1,z) = \wp(t;L)$. The only poles of $\wp(t)$ are double poles at the lattice points. To see that $\wp(t)$ is doubly periodic with periods given by the points of the lattice, first observe that its derivative $\wp'(t)$ is clearly doubly periodic. Then argue as in Koblitz [359, pp. 17–18].

The differential equation for $\wp(z)$ shows that it is the inverse function for the **elliptic integral** in Weierstrass normal form, i.e.,

$$z - z_0 = \int_{\wp(z_0)}^{\wp(z)} (4w^3 - g_2 w - g_3)^{-1/2} dw.$$

(see Siegel [596, Vol. I] and Lang [390]). The fact that the discriminant does not vanish on the upper half-plane shows that the cubic Eq. (3.63) has three distinct roots e_1, e_2, e_3 such that

Table 3.3 Ramanujan's tau function

n	1	2	3	4	5	6	7	8	9	10	11
$\boldsymbol{\tau(n)}$	1	−24	252	−1,472	4,830	−6,048	−16,744	84,480	−113,643	−115,920	534,612

$$\wp'(z)^2 = 4(\wp(z) - e_1)(\wp(z) - e_2)(\wp(z) - e_3).$$

One can show (see Apostol [12, p. 13] that the roots are given by values of $\wp$ at the half periods:

$$e_1 = \wp\left(\frac{1}{2}\right),\ e_2 = \wp\left(\frac{z}{2}\right),\ e_3 = \wp\left(\frac{1+z}{2}\right).$$

Then one can construct the function λ called the **modulus** by

$$\lambda(z) = (e_3 - e_2)/(e_1 - e_2). \tag{3.65}$$

It turns out that the function λ is an automorphic function for the congruence subgroup $\Gamma(2)$ (see Ahlfors [3, pp. 277–279]), and gives a useful conformal mapping. We will consider λ again later in this section.

The **Ramanujan $\boldsymbol{\tau}$-function** is defined on the positive integers as follows:

$$(2\pi)^{-12}\Delta(z) = \sum_{n\geq 1} \tau(n)\exp(2\pi i n z). \tag{3.66}$$

It is possible to use the Fourier expansion of the Eisenstein series given in Exercise 3.4.2 to see that the Ramanujan numbers $\tau(n)$ are all integers. Table 3.3 gives the first 11 of them.

The **Ramanujan conjecture** states that

$$|\tau(p)| < 2p^{11/2} \quad \text{for all primes } p. \tag{3.67}$$

This implies, using the multiplicative properties of the τ-function, that

$$|\tau(n)| \leq n^{11/2}\sigma_0(n) \qquad \text{where} \qquad \sigma_0(n) = \sum_{o<d|n} 1,\ \ n = 1,2,3,\ \ldots .$$

The multiplicative properties of the τ-function will be discussed in Sect. 3.6, using the theory of Hecke operators. Deligne proved the Ramanujan conjecture in the 1970s as well as its generalization to forms of weight k as a consequence of the Weil conjectures in algebraic geometry (see Deligne [126, 127] and the expository article of Katz [344, p. 14]). The generalization is called the Ramanujan–Petersson conjecture (see Theorem 3.4.4 and the discussion following it). There are also analogues for nonholomorphic Maass waveforms which have not been proved yet (see the discussion of the tables in Sect. 3.5 as well as that following Theorem 3.5.3 of Sect. 3.5).

Another question concerning $\tau(n)$ is still open: Can $\tau(n)$ vanish? The **Lehmer conjecture** says it cannot. In 1985 computer tables existed showing that $\tau(n) \neq 0$ for $n \leq 10^{15}$. J. Bosman has proved that $\tau(n) \neq 0$ for n less than a number close to 2×10^{19}. See Bosman's article in Couveignes and Edixhoven [112]. Still no one knows how to prove the Lehmer conjecture (see Lehmer [406–408]).

Write $\tau(p) = 2p^{11/2} \cos\varphi_p$. Inspired by the Sato–Tate conjecture for elliptic curves (see formula (3.78) below), Serre conjectured that the angles φ_p are distributed in $[0,\pi]$ with respect to the measure

$$\frac{2}{\pi}\sin^2\varphi \, d\varphi.$$

Lehmer [409] did a numerical study. Serre [580] and Ogg [507] showed that the Sato–Tate conjecture would follow from nonvanishing properties of the analytic continuation of L-functions associated to the mth symmetric power representation:

$$L_m(s) = \prod_{\substack{p\\ prime}} \prod_{i=0}^{m} (1 - \alpha_p^i \overline{\alpha_p}^{m-i} p^{-s})^{-1} \qquad \text{if} \begin{cases} \tau(p) = \alpha_p + \overline{\alpha_p},\ \alpha_p\overline{\alpha_p} = p^{11} \\ \alpha_p = p^{11/2}\exp(i\varphi_p). \end{cases}$$

The necessary results about $L_m(s)$ were proved for $m = 1$ by Hecke, for $m = 2$ by Rankin and Shimura, and for $m = 3,4$ by Jacquet and Shahidi. Additional references for this subject are Serre [579], Ram Murty [486, 487], Katz [345, p. 14], and Moreno and Shahidi [478]. Note that Lehmer's problem mentioned in the preceding paragraph is the question, Can $\varphi_p = \pi/2$? Nagoshi proved in 2006 that the Serre–Sato–Tate conjecture holds on average. See Nagoshi [489]. In 2011 the Serre–Sato–Tate conjecture was proved by Barnet-Lamb et al. [26]. The original Sato–Tate conjecture for elliptic curves has also been proved. See Clozel et al. [93].

Exercise 3.4.4. Another way to state the now proved Serre–Sato–Tate conjecture for the Ramanujan τ-function goes as follows. Let $\pi(x)$ be the number of primes $\leq x$, $a_p = \tau(p)p^{\frac{-11}{2}}$ and $[c,d] \subset [-2,2]$. The conjecture says that the a_p are distributed according to the semi-circle distribution, meaning that

$$\lim_{x\to\infty} \frac{1}{\pi(x)} \#\left\{p \leq x \;\middle|\; p \text{ prime},\ a_p \in [c,d]\right\} = \frac{1}{2\pi}\int_c^d \sqrt{1-x^2}\, dx.$$

Check that the histogram for a large set of a_p does resemble this semi-circle distribution.

As noted in Sect. 1.5, Wigner (see [738]) had the idea in the 1950s of making a toy model of the Schrödinger operator in quantum mechanics using large real symmetric $n \times n$ matrices with entries that are independent Gaussian random variables. He found that the histogram of the eigenvalues looks like a semi-circle (or, more precisely, a semi-ellipse). This has led people in physics to call the distribution of the preceding exercise the "Wigner semi-circle distribution"—around the same time that number theorists were calling it the "Sato–Tate distribution." Wigner's work was the beginning of a subject called "random matrix theory" and we will discuss it in Volume II [667].

Another useful fact about $\Delta(z)$ is **Jacobi's identity**:

$$\Delta(z) = (2\pi)^{12} q \prod_{n\geq 1} (1-q^n)^{24} \quad \text{for } q = \exp(2\pi i z). \tag{3.68}$$

Proofs of (3.68) can be found in Maass [437, p. 108], Ogg [505, pp. I-43–44], Rankin [532, pp. 196–197], Serre [576, pp. 95–96], and Weil [727, Chap. 4], for example.

The **eta function of Dedekind** is defined by

$$\eta(z) = q^{1/24} \prod_{n\geq 1} (1-q^n),\ q = \exp(2\pi i z). \tag{3.69}$$

Like the theta function in formula (3.50), the eta function is a modular form of weight $\frac{1}{2}$; but eta is a form for the whole modular group $SL(2,\mathbb{Z})$. In particular, eta satisfies

$$\eta(z+1) = \eta(z) \quad \text{and} \quad \eta(-1/z) = (z/i)^{1/2}\eta(z). \tag{3.70}$$

Note that, like theta, the eta function is an automorphic form with a multiplier system. If you can prove the last transformation formula in Eq. (3.70) for eta, then you can prove the product formula for the discriminant Eq. (3.68) since the product formulas imply that $\Delta = (2\pi)^{12}\eta^{24}$ and there is only one cusp form of weight 12. The transformation formula of eta is proved in many of the basic references on modular forms as well as in Pilar de la Torre and Goldstein [682] and Goldstein [224]. The complete multiplier system for eta involves Dedekind sums. We will have more to say about this in Sect. 3.7.7 on modular knots.

One of the main applications of the eta function to number theory comes from the fact that

$$q^{1/24}\eta(z)^{-1} = \sum_{n\geq 0} p(n) q^n,$$

where $p(n)$ is the number of **partitions** of n; i.e., the number of ways of writing

$$n = n_1 + \cdots + n_r,\ n_j \in \mathbb{Z},\ n_j > 0,\ j = 1,2,\ \ldots,\ r.$$

Rademacher used this fact to obtain an exact formula for $p(n)$ as an infinite series with terms involving certain finite trigonometric sums $A_k(n)$ and I-Bessel functions. The method is a further development of a technique of Hardy and Ramanujan from 1917, which begins with the Cauchy formula

$$p(n) = \frac{1}{2\pi i} \int_C x^{-n-1} f(x) dx,$$

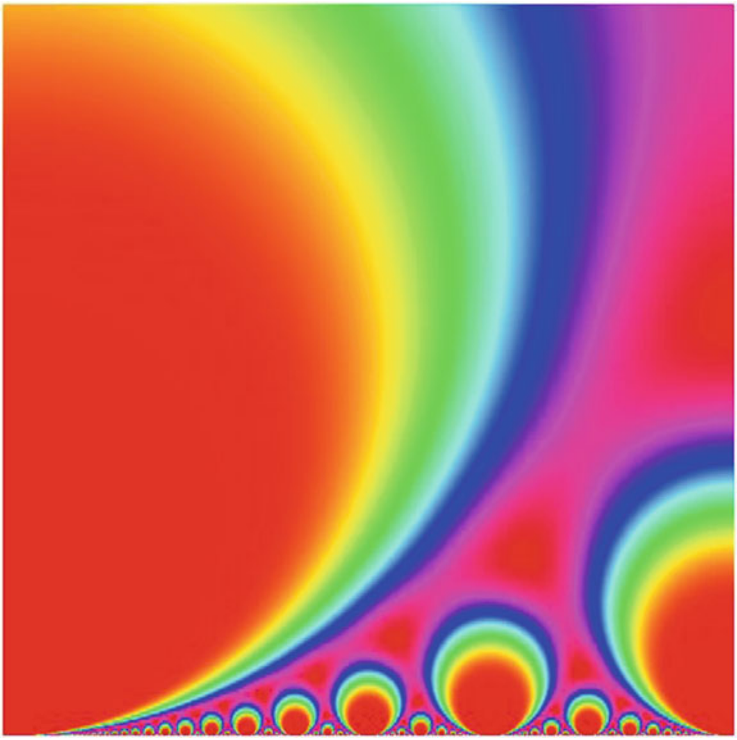


Fig. 3.33 A close up of $(2\pi)^{-12}y^6\,|\Delta(z)|$ produced in the same way as Fig. 3.30

with C a closed curve inside the unit circle and containing the origin. Here $f(q) = q^{1/24}\eta(z)^{-1}$. The path C is then taken as a circle around 0 and close to the unit circle. It is then cut into parts corresponding to neighborhoods of just one root of unity. For a discussion of the rather intricate arguments that ensue, see Rademacher [525, Chap. 14], or Lehner [410, p. 351]. The method is called the "circle method." For some history of other work on formulas for $p(n)$, see D.H. Lehmer's review of a paper of Whiteman which can be found in the collection of reviews in number theory edited by Leveque [414, Vol. IV, pp. 510–511] or by visiting the *Math. Reviews* website. Selberg (age 19 or so) obtained a formula for $A_k(n)$ as a finite Fourier series at around the same time that Rademacher had obtained his result. Lehmer had factored $A_k(n)$ to make computations.

The function $\eta(z)$ also appears in Kronecker's limit formula which has many applications in number theory (see Exercise 3.5.6 of Sect. 3.5). The eta function and its generalizations have also been used to obtain **class number formulas** (see Goldstein and Razar [226]).

Before leaving this section we look at Fig. 3.33, which is a close-up of $(2\pi)^{-12}y^6\,|\Delta(z)|$ produced in the same way as Fig. 3.30.

3.4.3 The Modular Invariant

Our next example of a modular form is the **modular invariant** $J(z)$ defined by

$$J = (60G_4)^3/\Delta. \tag{3.71}$$

This function was first constructed by Dedekind in 1877 and Klein in 1878. Here, once again, the definition seems mysterious, but we shall see that J is of fundamental importance, as is $j = 1728J$. The number 1728 gives j integer coefficients in its Fourier expansion. On the other hand, J has nice values at the elliptic and parabolic points (cusps) in a fundamental domain for $SL(2,\mathbb{Z})$.

Theorem 3.4.2 (The Main Properties of the Modular Invariant).

*(1) $J(z)$ is a **modular function**, i.e., an automorphic form of weight* 0*, which may have poles. J is holomorphic in H with a simple pole at ∞, $J(i) = 1$, and $J(\rho) = 0$, where $i = \sqrt{-1}$ and $\rho = (-1+\sqrt{-3})/2$.*
(2) J defines a conformal mapping which is one-to-one from D onto $\mathbb{C}$, where D is the fundamental domain for $SL(2,\mathbb{Z})$ in Fig. 3.14. Thus J provides an identification of $H/SL(2,\mathbb{Z}) \cup \{\infty\}$ with the Riemann sphere $\mathbb{C} \cup \{\infty\}$.
(3) The following are equivalent for a function f which is meromorphic on $H \cup \{\infty\}$:

   (a) f is a modular function (i.e., invariant under $SL(2,\mathbb{Z})$).
   (b) f is a quotient of two modular forms of the same weight.
   *(c) f is a rational function of J, (i.e., a quotient of polynomials in J) and thus J is called the **Hauptmodul** or fundamental function.*

Proof. (1) This property is clear from the properties of the discriminant and the Eisenstein series. The last two statements are proved by noting that $G_4(\rho) = 0$ and $G_6(i) = 0$, from formula (3.60). Thus $\Delta(i) = [60G_4(i)]^3$.
(2) To see this property, use formula (3.60) to show that $f_s = G_4^3 - s\Delta$ can only be zero at one point.
(3) To prove this property, first note that clearly (c)⇒(b)⇒(a). In order to prove (a)⇒(c), suppose that $f(z)$ is a modular function. Then multiply f by a polynomial in $J(z)$ to make a holomorphic function we will still call f on H. Then $g = \Delta^n f$ is holomorphic at infinity for some $n \geq 0$, to put g in $\mathcal{M}(SL(2,\mathbb{Z}), 12n)$. Thus g is a linear combination of $G_4^a G_6^b$, with $4a + 6b = 12n$. It suffices to show (c) for $f = G_4^a G_6^b/\Delta^n$. Now $p = a/3$ and $q = b/2$ are integers and $f = G_4^{3p} G_6^{2q}/\Delta^{p+q} = \left(G_4^3/\Delta\right)^p \left(G_6^2/\Delta\right)^q$. Finally, it is straightforward to show that G_4^3/Δ and G_6^2/Δ are rational functions of J, completing the proof.

□

Exercise 3.4.5. Fill in the details in the proof of Theorem 3.4.2.

Exercise 3.4.6. Produce some graphs of $J(z)$. Compare with Fig. 3.34.

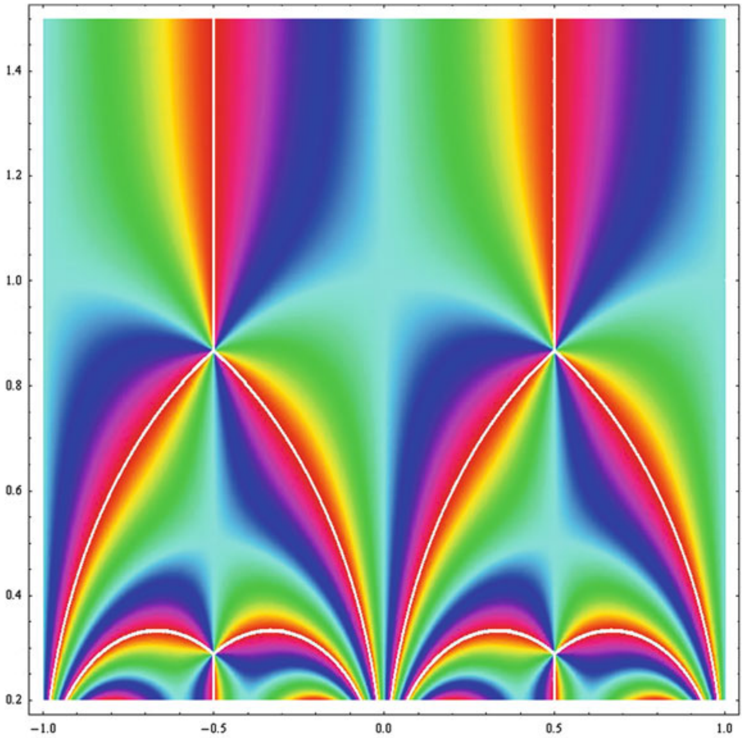


Fig. 3.34 A density plot of the argument of $J(z)$ produced with the Mathematica command: DensityPlot[Arg[KleinInvariantJ[x+I y]],{x,-1,1},{y,0.2,1.5}, ColorFunction->Hue, PlotPoints->400]

One can use the automorphic function J to prove the **small Picard theorem** very quickly. The inverse function $w = J^{-1}(z)$ is infinite valued with branch points at $0, 1, \infty$ (the values of J at the vertices of the fundamental domain). Suppose that f is an entire function omitting the values 0, 1, ∞. Then $J^{-1} \circ f = g$ is holomorphic in the z-plane and single valued. Since g lands in the upper half-plane, $\exp(ig)$ is bounded and thus constant by Liouville's theorem. It follows that f must be constant.

Still another, short proof of the little Picard theorem can be found in Kobayashi [358]. Moreover, Kobayashi's methods generalize to prove the *big* Picard theorem in higher-dimensional cases.

Another application of the J-function is in number theory. For the maximal unramified abelian extension of an imaginary quadratic number field is produced by a value of j, also called **the modular invariant**. An explicit reciprocity law shows how the Galois group of the extension is isomorphic to the ideal class group of the quadratic field. This is **explicit class field theory** (see our discussion of number fields in Sect. 1.4). In the ramified case, other functions than j are needed. A reference for this is Borel and Chowla et al. [53].

Martin Gardner wrote a famous April Fool's column for the April 1975 edition of *Scientific American* in which he claimed (among other bogus claims) that $e^{\pi\sqrt{163}}$ is an integer. Of course $e^{\pi\sqrt{163}}$ is transcendental by the theorem of Gelfand and Schneider (see Baker [21]). The fact is that $j\left(\frac{1+\sqrt{-163}}{2}\right) \in \mathbb{Z}$. The Fourier q-expansion of $j(z)$ in formula (3.72) implies then that $e^{\pi\sqrt{163}}$ is within 12 decimal places of an integer. Gardner noted later in the *Scientific American* blog: "To my amazement, thousands of readers failed to recognize the column as a joke."

It turns out that the Fourier or q-expansion of j has only integer coefficients:

$$j(z) = q^{-1} + 744 + 196884q + 21493760q^2 + \cdots, \tag{3.72}$$

where $q = \exp(2\pi iz)$. This is useful in the applications to number theory. There is also a surprising connection of the coefficients in Eq. (3.72) with the representations of the largest of the 26 sporadic finite simple groups which is known as the **Fischer-Griess monster group** M. The monster group has approximately 8×10^{53} elements. It happens that all of the early Fourier coefficients in Eq. (3.72) are simple linear combinations of degrees of representations of M. This was first noticed by John McKay and John Thompson. Some interesting analogues for Lie groups are explained by Kac [334]. See also Conway [105] and Conway and Norton [106]. Conway reported on the situation as follows:

> Because these new links are still almost completely unexplained, we refer to them collectively as the *moonshine* properties of the MONSTER, intending the word to convey our feelings that they are seen in a dim light, and that the whole subject is rather vaguely illicit!

In 1998, R. E. Borcherds won the Fields Medal in part because he proved the Conway–Norton "moonshine conjecture." See Borcherds [49].

The modular invariant j is important to algebraic geometers because it characterizes elliptic curves (see Shimura [589, p. 97]). We will say more about elliptic curves later in this section when we discuss elliptic integrals. The reason that the modular group is so called is that the moduli variety classifying elliptic curves is $SL(2,\mathbb{Z})\backslash H$.

3.4.4 Theta Functions

The theta function is an important example for us (like G_k and unlike J and Δ) because generalizations to higher dimensions are known. The **theta function associated to a positive definite symmetric $n \times n$ real matrix P and a point z in the upper half-plane** is

$$\theta(P,z) = \sum_{a\in\mathbb{Z}^n} \exp(\pi i P[a] z). \tag{3.73}$$

Here $P[a]$ denotes the quadratic form associated to P; i.e., $P[a] = {}^t aPa$, thinking of a as a column vector in $\mathbb{Z}^n$, as usual. Note that if $z = it$, for $t > 0$, them $\theta(P,z) = \theta(P,t)$ as defined in Exercise 1.4.6 of Sect. 1.4. Note also that if $n = 1,\ P = 1,\ \theta(1,z) = \theta(z)$, the function defined by formula (3.50) in Sect. 3.3.

Exercise 3.4.7. Prove the **transformation formula**

$$\theta(P,z) = |P|^{-1/2} \left(\frac{z}{i}\right)^{-n/2} \theta(P^{-1}, -1/z).$$

Hint. *Recall Exercise 1.4.6 of Sect. 1.4.*

In order for the theta function defined by Eq. (3.73) to be a modular form as a function of z (with trivial multiplier system for the group $SL(2,\mathbb{Z})$), we need to place restrictions on P.

Theorem 3.4.3. *We say the positive definite symmetric $n \times n$ real matrix P* ***is even integral*** *if $P[a] \in 2\mathbb{Z}$ for all $a \in \mathbb{Z}^n$. If $\theta(P,z)$ is defined by formula (3.73) and P is even integral of determinant 1, then 8 divides n and $\theta(P,z)$ is in $\mathcal{M}(SL(2,\mathbb{Z}), n/2)$.*

Proof. First we need an exercise.

Exercise 3.4.8. Show that under the hypotheses on P in Theorem 3.4.3, the size n of the matrix P must be divisible by 8.

Hint. *(See Serre [576, pp. 53 and 109].) You can assume that $n \equiv 4 (\mathrm{mod}\, 8)$ by replacing P by*

$$\begin{pmatrix} P & 0 \\ 0 & P \end{pmatrix} \quad \text{or} \quad \begin{pmatrix} P & 0 & 0 & 0 \\ 0 & P & 0 & 0 \\ 0 & 0 & P & 0 \\ 0 & 0 & 0 & P \end{pmatrix}.$$

Then let $V = ST$, with $Sz = -1/z$ and $Tz = z + 1$. Compute $\theta(P, V(iy))$ for $y > 0$, as well as $\theta(P, V^3(iy))$. Obtain a contradiction from $V^3 = I$.

From Exercises 3.4.7 and 3.4.8, it follows that $\theta(P, z+1) = \theta(P,z)$ and $\theta(P,z) = z^{-n/2}\theta(P, -1/z)$, if P satisfies the hypotheses of Theorem 3.4.3. For if $|P| = 1$ and P is integral, then, upon setting

$$S_P = \{P[a] \mid a \in \mathbb{Z}^n\},$$

we see that $S_P = S_{P^{-1}}$. This is derived from the fact that we can write $P^{-1}[b] = P[a]$ for $b = P[a] \in \mathbb{Z}^n$. □

We call two positive definite real $n \times n$ matrices P and Q **equivalent modulo $GL(n,\mathbb{Z})$** if there is a matrix A in $GL(n,\mathbb{Z})$ = the integral $n \times n$ matrices of determinant $+1$ or -1, such that $P = Q[A] = {}^t AQA$. In fact, this equivalence relation will be quite important in Volume II [667] as an analogue of the equivalence relation defining the quotient space $SL(2,\mathbb{Z})\backslash H$.

Exercise 3.4.9. Show that the following 8×8 matrix, which comes out of the theory of the exceptional Lie group E_8, gives an example of a matrix satisfying the hypotheses of Theorem 3.4.3. It can, in fact, be proved that this is the only possible example, up to equivalence modulo $GL(8,\mathbb{Z})$.

$$\begin{pmatrix} 2 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & -1 & 0 & 0 & 0 & 0 \\ -1 & 0 & 2 & -1 & 0 & 0 & 0 & 0 \\ 0 & -1 & -1 & 2 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 2 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 2 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & 2 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 2 \end{pmatrix}.$$

For $n = 16$, there are two examples of matrices satisfying the hypotheses of Theorem 3.4.3 and nonequivalent modulo $GL(16,\mathbb{Z})$. For $n = 24$, there are 24 examples nonequivalent modulo $GL(24,\mathbb{Z})$ (see Serre [576, Chaps. 5 and 7]).

If one considers relaxing the hypotheses on P in Theorem 3.4.3; e.g., allowing P to have arbitrary determinant, then one obtains a modular form attached to a **congruence subgroup**

$$\Gamma_0(N) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL(2,\mathbb{Z}) \,\middle|\, N \text{ divides } c \right\}.$$

The multiplier system for such theta functions will not be trivial. The theory is harder for odd n, even if $n = 1$, which is the most widely used special case—the theta function corresponding to the Riemann zeta function by Mellin transform. The **level** N of the congruence subgroup is called the level of the quadratic form P and defined to be the least positive integer such that NP^{-1} is even integral. For n even, it was proved by Hermite that

$$\theta(P,\gamma z) = \varepsilon(d)(cz+d)^k \theta(P,z) \text{ for } \gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma_0(N),\ k = n/2,$$

where the multiplier system is given by

$$\varepsilon(d) = (\operatorname{sign} d)^k \left(\frac{(-1)^k |P|}{|d|} \right). \tag{3.74}$$

The symbol $(-)$ in the multiplier system is the **Kronecker symbol**, which is a generalization of the Legendre symbol. The **Legendre symbol** $\left(\frac{a}{p}\right)$ is defined, for $a,\ p$ integers with p an odd prime not dividing a, by

$$\left(\frac{a}{p}\right) = \begin{cases} +1 & \text{if } a \equiv x^2 (\operatorname{mod} p) \ \text{ has a solution } \ x \in \mathbb{Z}, \\ -1 & \text{otherwise.} \end{cases} \tag{3.75}$$

The definition of the Kronecker symbol can be found in most elementary number theory texts, and, for example, in Cohn [97], Grosswald [233], or Siegel [598].

The **quadratic reciprocity law**, first proved by Gauss (eight times), says that if p and q are distinct odd primes then

$$\left(\frac{p}{q}\right)\left(\frac{q}{p}\right) = (-1)^{\frac{p-1}{2} \cdot \frac{q-1}{2}}. \tag{3.76}$$

It is possible to prove the quadratic reciprocity law (one of the most basic truths in number theory) using theta functions (see the next exercise). Hecke generalized this argument to obtain the quadratic reciprocity law for algebraic number fields, using theta functions (see Hecke [260] and Siegel [600, Vol. III, Paper 74]). Proofs of formula (3.74) can be found in Eicher [150, pp. 48–52], Ogg [505, Chap. 9] and Schoeneberg [562, Chap. 9]. The proof given by Eichler is quite interesting because it uses theta functions on the Siegel upper half-space (see Volume II [667] for an introduction to such Siegel modular forms). Moreover, Eichler's proof works for an odd number of variables. The latter case is due to Pfetzer [513]. There is also a very interesting article by Eichler [151], which includes another proof of the result for an even number of variables plus some history and connections with other areas; e.g., the representations of $SL(2, \mathbb{Z}/n\mathbb{Z})$. Actually, one must generalize the theta functions defined in this section and introduce spherical harmonics in the sums (the same spherical harmonics that we considered in Sect. 2.1) in order to obtain cusp forms and bases for spaces of modular forms for $\Gamma_0(N)$ (see also Hijikata et al. [295], Serre and Stark [581]). For other discussions of theta functions and multiplier systems, see Knopp [357], Maass [437], Mumford [485], and Rankin [532].

Exercise 3.4.10. Prove the following **formula of Landsberg and Schaar**:

$$\frac{1}{\sqrt{p}} \sum_{n=0}^{p-1} \exp\left(2\pi i n^2 \frac{q}{p}\right) = \frac{e^{in/4}}{\sqrt{2q}} \sum_{n=0}^{2q-1} \exp\left(\frac{-i\pi n^2 p}{2q}\right),$$

for any integers p and q.

Hint. *Use the transformation formula for theta (for $z \in H$):*

$$\sum_{n \in \mathbb{Z}} \exp(i\pi n^2 z) = (z/i)^{-1/2} \sum_{n \in \mathbb{Z}} \exp(i\pi n^2(-1/z)).$$

Let $z = 2q/p + it$ and let t approach zero. It is possible to use this exercise to prove the quadratic reciprocity law (see Dym and McKean [147, pp. 222–226], and Bellman [33, Sect. 29]).

For the applications to elliptic integrals, it is useful to define slightly more general theta functions, similar to that in Exercise 1.3.8 of Sect. 1.3. These functions will be studied in the following pair of exercises. There are multitudes of theta functions, and the notation for these functions is definitely not standardized. So be careful if you consult references. Here are some books to consult, in addition to our standard list from the beginning of Sect. 3.3 and the books just mentioned: Bellman [33], Erdélyi et al. [164], Hurwitz and Courant [312], Igusa [313], Lion and Vergne [423], Rauch and Lebowitz [536], and Whittaker and Watson [734]. Of course, some of these theta functions are known to Mathematica. For example, the Mathematica command `EllipticTheta[a,u,q]` gives the Jacobi theta function $\theta_a(u,q)$ in Exercise 3.4.12.

Exercise 3.4.11 (Another Theta Function). Define

$$\vartheta(z,s) = \sum_{n\in\mathbb{Z}} exp\left(\pi i z n^2 + 2\pi i s n\right), \quad \text{for } z \in H \quad \text{and} \quad s \in \mathbb{C}.$$

Prove the following statements about this theta function.

(a) $\vartheta(z,s)$ is an **entire** function of $z \in H$ and of $s \in \mathbb{C}$.
(b) $\vartheta(z,s)$ is **quasi-periodic** as a function of s, in the following sense:

$$\vartheta(z,s+n) = \vartheta(z,s), \text{ for all } n \in \mathbb{Z};$$
$$\vartheta(z,s+zn) = \exp\left[-\pi i\left(zn^2+2ns\right)\right]\vartheta(z,s), \text{ for all } n \in \mathbb{Z}.$$

(c) $\vartheta(z,s)$ satisfies the **transformation formula**

$$\vartheta(z,s) = (z/i)^{-1/2}\sum_{n\in\mathbb{Z}}\exp\left(\frac{-\pi i\,(n-s)^2}{z}\right).$$

(d) $\vartheta\left(z,(1+z)/2\right) = 0$.
(e) The only **zero** of $\vartheta(s) = \vartheta(z,s)$ as a function of s, holding z fixed, for s in the period parallelogram on 1 and z is $s = (1+z)/2$. Moreover, this zero is simple.

Hints. For (c) recall Poisson summation. For (e) imitate the proof of formula (3.60). This theta function is considered by Siegel [596, Vol. II, pp. 158–172] and Mumford [485, Vol. I] along with generalizations to the Siegel upper half-plane.

Exercise 3.4.12 (Jacobi's Four Theta Functions). Define for $q = exp(\pi i z)$, $z \in H$,

$$\theta_1(u,q) = 2\sum_{n=0}^{\infty}(-1)^n q^{(n+\frac{1}{2})^2}\sin[(2n+1)u];$$

$$\theta_2(u,q) = 2\sum_{n=0}^{\infty} q^{(n+\frac{1}{2})^2}\cos[(2n+1)u];$$

$$\theta_3(u,q) = 1+2\sum_{n=1}^{\infty} q^{n^2}\cos[2nu];$$

$$\theta_4(u,q) = 1+2\sum_{n=1}^{\infty} (-1)^n q^{n^2}\cos[2nu].$$

Prove that

$$\theta_3(u,q) = \vartheta\left(z,\frac{u}{\pi}\right);\ \theta_4(u,q) = \vartheta\left(z,\frac{u}{\pi}+\frac{1}{2}\right);$$

$$\theta_2(u,q) = \theta_1\left(u+\frac{\pi}{2},q\right);\ \theta_1(u,q) = -i\exp\left(iu+\frac{\pi iz}{4}\right)\theta_4\left(u+\frac{\pi z}{2},q\right).$$

We will show later that you can use products and quotients of the theta functions in Exercise 3.4.11 to obtain any elliptic function (see the discussion of elliptic integrals which follows). This aids in the computation of elliptic functions, because the series for theta functions converge rapidly, when the variable z is away from the real line. And you can use the transformation formula in part (c) of Exercise 3.4.11 to deal with other values of z. Siegel extends these results to generalizations of elliptic functions called Jacobi-Abel functions in [596], using Siegel modular forms to be considered in Volume II [667]. The Jacobi-Abel functions are related to abelian integrals (see Sect. 3.4.5).

Set

$$\vartheta_{11}(z,s) = \exp\left[\frac{1}{4}\pi iz+\pi i(s+\frac{1}{2})\right]\vartheta\left(z,s+\frac{1}{2}(1+z)\right).$$

Mumford [485, Vol. I, p. 25] shows that, up to a constant, the Weierstrass $\wp$-function is obtained from $\vartheta_{11}(z,s)$ by taking the logarithm of theta and then two derivatives in the s-variable. We should caution the reader that our notation is the mirror image of Mumford's. Mumford [loc. cit., p. 64] proves Jacobi's derivative formula which shows that taking derivatives of $\vartheta_{11}(z,s)$ with respect to s and then evaluating at $s=0$ leads to cusp forms of weight 3 and level 4 (after squaring the result). And Mumford [loc. cit., p. 72] shows that $\Delta(z)$ can be written as a 24th power of a value of $\vartheta(z,s)$.

Theta functions are viewed by algebraic geometers as providing "nice projective embeddings of a polarized abelian variety" (see Igusa [313] and Mumford [485]). There are connections between this view and the application to elliptic integrals which follows. A historical sketch explaining some of these things can be found in Shafarevitch [582, pp. 411–430 and Chap. 11]. Another reference is Hoobler and Resnikoff [300]. Much of the modern algebraic-geometric theory of theta functions (of higher genus) has been developed by Mumford (see Tate [657] and Mumford [482–485]). The higher-genus theta functions will be considered in Volume II [667]. The book of Mumford [485] is recommended as a nice introduction to theta functions which includes several applications to the solution of partial differential equations.

There are numerous applications of theta functions in number theory; e.g., in order to obtain correspondences between various spaces of automorphic forms (see Kudla [377, 378], Stark [628], and Marie-France Vignéras [699]), to study representations of integers by quadratic forms and related questions (see Exercises 3.4.19, 3.4.20, and Siegel [600, Vol. I, pp. 326–405, 410–443, 469–548; Vol. II, pp. 1–7, 20–40]). Siegel's work on quadratic forms finds a new adelic and representation-theoretic formulation in Weil [729, Vol. III, pp. 1–157] and Tamagawa (see Kneser's article in Cassels and Fröhlich [79]).

3.4.5 Elliptic Integrals and Elliptic Curves

The name "elliptic integral" arises because the arc length of an ellipse is such an integral. An **elliptic integral** is an integral of the form $\int R(x,y)dx$, where $R(x,y)$ is a rational function over $\mathbb{C}$ (i.e., a quotient of polynomials with complex coefficients) and x and y are related by an algebraic equation which can be solved by substitutions $x = f(u)$ and $y = g(u)$, with f, g being elliptic functions; e.g., $y^2 = G(x)$, where $G(x)$ is a third- or fourth-degree polynomial over $\mathbb{C}$. Here an **elliptic function** means a meromorphic function of $z \in \mathbb{C}$ which is doubly periodic; i.e., has two linearly independent periods over $\mathbb{R}$. The best example is the Weierstrass $\wp$-function of formula (3.64).

Abelian integrals are generalizations of elliptic integrals, where the equation relating x and y is of higher degree. If the equation is $y^2 = G(x)$, where $G(x)$ has degree greater than 4, the integral is **hyperelliptic**.

Elliptic functions arise in many ways in applied mathematics; e.g., in potential theory on rectangles, in problems involving pendulums, and in problems involving ellipses (see Jeffreys and Jeffreys [327, Chap. 25]). Perhaps one of the strangest things about elliptic functions is that, unlike the well-known functions of freshman mathematics, they satisfy **nonlinear differential equations** (e.g., see formula (3.63) for the Weierstrass $\wp$-function, Exercises 3.4.14 and 3.4.21).

Elliptic integrals arise, for example, when one uses the Schwarz–Christoffel formula to obtain the conformal mapping of the rectangle in Fig. 3.35 onto the upper half-plane. Here the points are mapped according to Table 3.4 and $sn'(0) = 1$.

The integral defining the inverse function for the **Jacobian elliptic function** $w = sn(z,k) = sn(z)$ is

$$z = \int_0^w [(1-w^2)(1-k^2w^2)]^{-1/2} dw. \tag{3.77}$$

Note that the case $k = 0$ gives the ordinary trigonometric function $w = \sin z$.

The Jacobian elliptic function is a doubly periodic function with periods $4K$ and $2iK'$. The **modulus** k^2 (as well as k with $0 < k < 1$) is uniquely determined by $q = \exp(i\pi\tau)$, $\tau = iK/K'$. In the theory of elliptic integrals, k^2 was always called the modulus and it turns out that

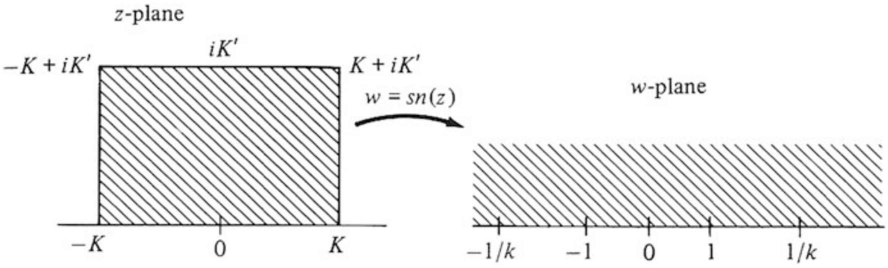


Fig. 3.35 Conformal mapping by $sn(z)$

Table 3.4 How points in the rectangle of Fig. 3.35 are mapped by the Jacobian elliptic function

z	$-K+iK'$	$-K$	0	K	$K+iK'$	iK
$sn(z)$	$-k^{-1}$	-1	0	-1	k^{-1}	∞

$$k^2(q) = \lambda(\tau),$$

where λ was defined in formula (3.65). This explains why we call λ the "modulus." References for these things are Copson [109, Chap. 14], Du Val [146], and Nehari [492, pp. 280–296], for example.

The elliptic function $sn(z)$ is used in applied mathematics to produce the potential distribution inside a rectangle when there is some kind of charge distribution on the boundary (see Morse and Feshbach [479, Vol. II, p. 1252]). Reading this page in Morse and Feshbach, one is struck by the regret with which they turn to higher-dimensional problems from those of the complex plane, for higher-dimensional problems seldom have such elegant solutions. The method of conformal mapping mostly fails. For Liouville proved that the only nontrivial conformal maps in 3-space are "mappings by reciprocal radii" (inversions in spheres)—a fact also bemoaned by Sommerfeld [610, p. 140].

Exercise 3.4.13. Use the Schwarz–Christoffel transformation to show that the Jacobian elliptic function has the properties that we claimed above.

Exercise 3.4.14. Show that $w = sn\,(z)$ satisfies

$$\left(w'\right)^2 = (1-w^2)(1-k^2w^2).$$

Exercise 3.4.15. Show that if $q = e^{\pi i \tau}$,

$$\frac{\theta_1(z,q)}{\theta_4(z,q)} = A\ sn\left(\frac{2K}{\pi}z\right).$$

Hint. *Compare zeros and poles of the two doubly periodic functions. The functions* θ_1 *and* θ_4 *are defined in Exercise 3.4.12.*

Exercise 3.4.16. Show that the constant A in Exercise 3.4.15 has the value

$$A = \frac{\theta_2(0,q)}{\theta_3(0,q)}.$$

Also show that

$$\lambda = k^2 = \left(\frac{\theta_2(0,q)}{\theta_3(0,q)}\right)^4.$$

Since the theta series converge rapidly, these formulas can be used to evaluate the functions above on a computer.

There are many other known and useful formulas for these functions. We mention, for example, the **relation between the modulus and *J***:

$$J = \frac{4}{27}\frac{(1-\lambda+\lambda^2)^3}{\lambda^2(1-\lambda)^2}.$$

Jacobi's triple product formula says that if $z = \exp(2iu)$

$$\theta_3(u,q) = \prod_{n=1}^{\infty}(1\ -q^{2n})(1+q^{2n-1}z^2)(1+q^{2n-1}z^{-2}).$$

For a proof, see Hurwitz and Courant [312], for example. This and related identities for the Dedekind eta function have attracted much attention recently thanks to connections with Lie groups (see Kac [335] and MacDonald [441]). For more connections between theta functions and the theory of representations of Lie groups, see Lion and Michelle Vergne [423] and Mumford [485].

The words "elliptic curve" have arisen many times in this section. Let's now seek to understand a bit of the basic facts about them. The most concrete way to define an **elliptic curve** over $\mathbb{C}$ is as the points $(x,y) \in \mathbb{C}^2$ satisfying the equation $y^2 = ax^3 + bx^2 + cx + d$ provided that the discriminant of the cubic polynomial on the right is non-zero. The equation is said to be in Weierstrass form if it is $y^3 = 4x^2 - g_2x - g_3$ [as in formula (3.63)]. Suppose that L is the lattice $\mathbb{Z} \oplus \tau\mathbb{Z}$ and $\wp(z) = \wp(z,\tau) = \wp(z,L)$, with the Weierstrass function $\wp(z)$ defined by formula (3.64). Then the mapping that sends $z \in \mathbb{C}^2/L$ to $(\wp(z), \wp'(z))$ gives an identification of the torus $\mathbb{C}^2/L$ with points on the curve $y^3 = 4x^2 - g_2x - g_3$. This allows one to add points on this curve by adding the corresponding points on the torus. That leads to the group of the curve. There is also a way to add points on a curve directly which is useful if you want to replace $\mathbb{C}$ by a finite field. The curve will have to have a well-defined tangent at every point since to add a point P to itself you need to intersect the tangent at P with the curve. The curve in Fig. 3.36 illustrates how to add point A to point B over the reals.

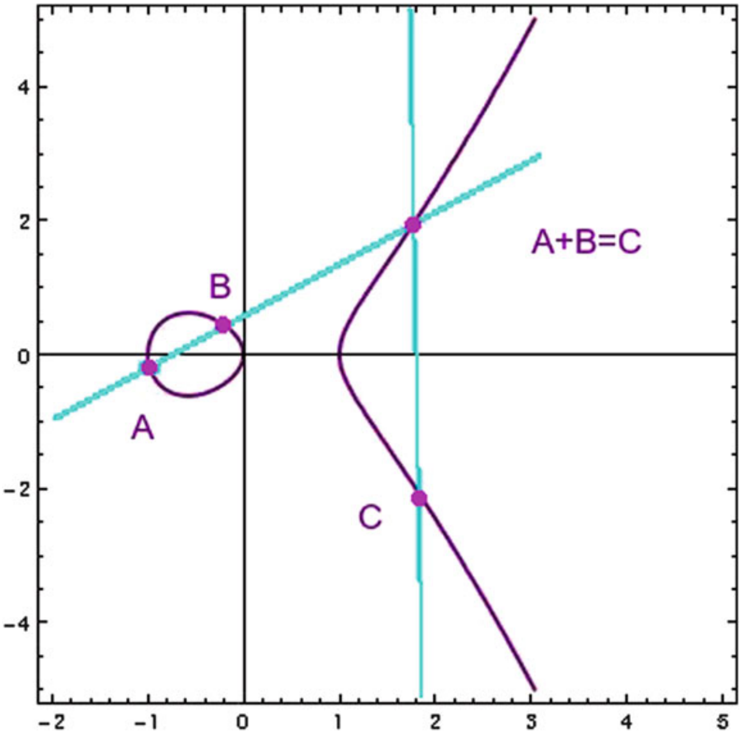


Fig. 3.36 A picture of the addition $A+B=C$ on the elliptic curve $y^2=x^3-x$ where the field is $\mathbb{R}$

Once you have defined elliptic curves over $\mathbb{C}$ or $\mathbb{R}$ or $\mathbb{Q}$, you can replace these fields by finite fields. Then it is possible to apply elliptic curves to cryptography. See Koblitz [360]. It is useful then to know some basic theorems about this case. For example, Hasse's theorem says that if p is a prime and N_p is the number of points on an elliptic curve over the finite field $\mathbb{F}_p$, then $\left|N_p-(p+1)\right|\leq 2\sqrt{p}$. Thus we can write $p+1-N_p=2\sqrt{p}\cos\theta_p$. Independently around 1960 Mikio Sato and John Tate made a conjecture. The **Sato–Tate conjecture for elliptic curves without complex multiplication** says:

$$\lim_{x\to\infty}\frac{\#\{p\text{ prime}\mid p\leq x\text{ and }a\leq\theta_p\leq b\ \}}{\#\{p\text{ prime}\mid p\leq x\ \}}=\frac{2}{\pi}\int_a^b\sin^2 u\,du. \tag{3.78}$$

This has been proved by Clozel et al. [93].

The popular groups for public-key cryptography using the RSA system were once the multiplicative groups of units (or invertible elements for multiplication) in the ring $\mathbb{Z}/pq\mathbb{Z}$, where p and q are very large primes. The difficulty of factoring the now very very large pq, makes the discrete logarithm problem in this group very very hard. However, in 1985 Neil Koblitz and Victor Miller proposed cryptosystems based on elliptic curves. The groups of elliptic curves are the groups behind much of the encryption in use at the moment. They have the advantage of being much smaller than the groups needed for RSA cryptography. They make up much of the protection for your passwords on your Internet transactions. See Kristin Lauter's article [397] for a brief introduction to the state of the art as of 2004. Koblitz [361] discussed the "drama and conflict ... inherent in cryptography, which, in fact, can be defined as the science of transmitting and managing information in the presence of an adversary." Koblitz and Menezes [362] discuss mistakes in "proofs of security" of various cryptosystems and why, despite this, we should not worry about the security of internet transactions.

3.4.6 The Connection Between Theta Functions and Coding Theory

References for this subject are Sloane [606, 607], Jessie MacWilliams and Sloane [445], and Conway et al. [107]. The codes discussed here are the error-correcting codes, not the secret ones mentioned in the last paragraph on elliptic curve cryptography. Much of the foundations of error-correcting codes was developed at AT&T Bell labs. We shall consider only binary codes. These are vector spaces over the field $\mathbb{F}_2$ with two elements.

Consider a noisy telephone connection transmitting 0s and 1s. Sometimes a 1 is transmitted but is received as a 0. Thus the messages need to be encoded to include redundancy in order to correct errors. The coding theorist corrects errors by sending messages encoded into **codewords** which are vectors of 0s and 1s. The **Hamming distance** between two codewords x, y is the number of indices i in the vectors $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$ such that $x_i \neq y_i$.

Definition 3.4.2. A **binary code** C of type $[n,k,d]$ is a set of 2^k vectors $u = (u_1, \ldots, u_n)$ with $u_i \in \mathbb{Z}/2\mathbb{Z} = \mathbb{F}_2$ for all $i = 1, \ldots, n$ such that both of the following hold:

(a) C is closed under addition (componentwise and mod 2).
(b) The Hamming distance between two codewords is $\geq d$.

Coding theorists have the incompatible goals of seeking n to be small for speed of transmission, k to be large for efficiency, and d to be large to correct many errors. It turns out that a code $[n,k,d]$ can correct $\lfloor (d-1)/2 \rfloor$ errors.

Suppose that C is a **linear** code of type $[n,k,d]$; i.e., C forms a k-dimensional vector subspace of $\mathbb{F}_2^n$ with scalars given by elements of the finite field $\mathbb{F}_2$. Then it is shown in Sloane [606] that one can obtain a lattice in $\mathbb{R}^n$ from C in several ways (see Sect. 1.4 for the definition of lattice and the connection with positive definite symmetric matrices). One way of obtaining a lattice from C (and the only one that we shall discuss) is to set

$$L(C) = \left\{ (u+2m)/\sqrt{2} \;\middle|\; u \in C, m \in \mathbb{Z}^n \right\}.$$

Here we think of u as a vector of 0s and 1s.

For example, the code $C = \{000, 011, 101, 110\}$ of type $[3,2,2]$ gives rise to the face-centered cubic lattice (see Sect. 1.4) in this way and thus to the densest lattice packing of spheres in $\mathbb{R}^3$ (see Volume II which is [667]).

If $u \in C$, the **weight** of $u = wt(u)$ is defined to be the number of components u_i of u such that $u_i \neq 0$. And the **weight enumerator** of C is a homogeneous polynomial of degree n in two indeterminants x and y defined by

$$w_C(x,y) = \sum_{u \in C} x^{n-wt(u)} y^{wt(u)}.$$

It has been proved by Berlekamp, Jessie MacWilliams, and Sloane as well as Broue and Enguehard (see Sloane [606]) that defining the **theta function** associated to $L(C)$ by

$$\theta_{L(C)}(z) = \sum_{a \in L(C)} \exp(\pi i z\, {}^t aa), \quad z \in H,$$

then $\theta_{L(C)}$ is actually a weight enumerator with Jacobi theta functions from Exercise 3.4.12 substituted for x and y. The explicit formula is

$$\theta_{L(C)}(z) = w_C(\theta_3(0,q), \theta_2(0,q)) \quad \text{if} \quad q = e^{2\pi i z}.$$

In particular, it is possible to use a code called the extended Hamming code H_8 to obtain the theta function mentioned earlier, which is associated to the Lie group E_8 (see Exercise 3.4.9). To obtain the Hamming code H_8, one writes down a matrix called the parity check matrix:

$$H = \begin{pmatrix} 1&1&1&1&1&1&1&1 \\ 0&0&0&1&1&1&1&0 \\ 0&1&1&0&0&1&1&0 \\ 1&0&1&0&1&0&1&0 \end{pmatrix}.$$

Then the code words are obtained by choosing components u_1, u_2, u_3, u_4 arbitrarily in $\mathbb{F}_2$ while choosing the remaining components of u so that the equation $Hu = 0$ holds (modulo 2, of course). This is 4 equations in 4 unknowns, so that the solution is unique. And H_8 is a code of type $[8,4,4]$.

3.4.7 Poincaré Series

There is another example of a modular form which has been useful—the **Poincaré series**, which generalizes the Eisenstein series G_k defined in formula (3.59) by replacing

$$(cz+d)^{-k} = \left(\frac{d(\gamma z)}{dz}\right)^{k/2} \quad \text{for } \gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL(2,\mathbb{Z}),$$

by

$$f(\gamma z)(cz+d)^{-k},$$

for some holomorphic bounded function $f : H \to \mathbb{C}$. Usually, a function such as $f(z) = \exp(2\pi imz)$ for $m \in \mathbb{Z}$ is chosen (see Lehner [410, p. 275] and Rankin [532, p. 136]). Petersson showed that you can obtain a basis for $\mathcal{M}(SL(2,\mathbb{Z}),k)$ of such Poincaré series (see Lehner [410, p. 293] and Rankin [532, pp. 153–159]).

These Poincaré series are also useful in the study of automorphic forms for congruence subgroups (see Rankin [532, pp. 172–193]). In his article on the basis problem for modular forms on $\Gamma_0(N)$, Eichler uses the Poincaré series with $f_w(\gamma,z) = \chi(\gamma)(\gamma(z)-w)^{-1}$, where $\chi(\gamma) = \chi(a)$, for χ a character of the group $(\mathbb{Z}/N\mathbb{Z})^* =$ the multiplicative group of integers a mod N with $(a,N) = 1$ (see Eichler [151, p. 122]). These Poincaré series are used to construct a Green's function to be used in the trace formula. The Fourier coefficients of Poincaré series involve interesting arithmetical sums called Kloosterman sums (see formula (3.97) and Rankin [532]). We should note that one of the big problems with Poincaré series is that **they may be identically zero** (see Lehner [410, p. 25, pp. 290–291]).

Finally, we should mention that Poincaré series (like Eisenstein series) are capable of extensive generalization. For example, Svetlana Katok [342] considers **relative Poincaré series attached to hyperbolic elements** $\gamma_0 \in \Gamma$, where Γ is a Fuchsian group of the first kind with compact fundamental domain. Equivalently, one can say that these series are attached to closed or periodic geodesics in the fundamental domain of Γ. She proves that these series generate the whole space of cusp forms of weight greater than or equal to 4. These series are sums over $\Gamma_0\backslash\Gamma$, where Γ_0 is the **centralizer** in Γ of γ_0; i.e., Γ_0 is the set of elements of Γ commuting with γ_0. If

$$\gamma_0 = \begin{pmatrix} a_0 & b_0 \\ c_0 & d_0 \end{pmatrix},$$

set

$$Q_{\gamma_0}(z) = c_0 z^2 + (d_0 - a_0)z - b_0.$$

Then the relative Poincaré series for γ_0 is essentially the sum over γ representing $\Gamma_0\backslash\Gamma$ of

$$Q_{\gamma_0}(\gamma z)^{-k/2}(cz+d)^{-k} \quad \text{where } \gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

The construction works for groups with cusps also (see Exercise 3.4.17). Furthermore, Katok shows that the periods over closed geodesics play as important a role as the Fourier coefficients play for the group $SL(2,\mathbb{Z})$ and congruence subgroups.

Exercise 3.4.17 (from Svetlana Katok [342]) (Relative Poincaré Series for SL(2, $\mathbb{Z}$)). Let $\gamma_0 \in SL(2,\mathbb{Z})$ be a hyperbolic element. With the notation above, consider the relative Poincaré series defined by:

$$\sum_{\Gamma_0\backslash\Gamma} Q_{\gamma_0}(\gamma z)^{-k/2}(cz+d)^{-k}, \quad \text{where} \quad \gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL(2,\mathbb{Z}).$$

Show that the series represents a modular form of weight k for $k \geq 4$.

3.4.8 Some Estimates for Fourier Coefficients of Modular Forms

We have noted previously the Ramanujan conjecture bounding Fourier coefficients of certain cusp forms for $SL(2,\mathbb{Z})$. Here we look at non-cusp forms and we prove a weaker result for cusp forms.

Theorem 3.4.4 (Estimates for the Fourier Coefficients of Modular Forms). *The Fourier coefficients of the Eisenstein series $G_k(z)$ defined by Eq. (3.59)*

$$G_k(z) = \sum_{n\geq 0} a_n exp(2\pi inz), \qquad (k = 4, 6, 8, 10, 12, \ldots)$$

satisfy

$$An^{k-1} \leq a_n \leq Bn^{k-1} \quad \text{for some positive constants } A, B.$$

The Fourier coefficients of a cusp form $f(z)$ of weight k for $SL(2,\mathbb{Z})$ with

$$f(z) = \sum_{n\geq 1} a_n exp(2\pi inz)$$

satisfy the ***Ramanujan–Petersson bound*** *(proved by Deligne [126, 127]):*

$$|a_n| \leq Cn^{\frac{k-1}{2}+\varepsilon}, \qquad \text{for any} \quad \varepsilon > 0.$$

Proof (of a weaker result). To prove the estimate for the coefficients of G_k, note that Exercise 3.4.2 implies that

$$a_n = C\sigma_{k-1}(n)$$

for some constant $C \neq 0$, where σ_s denotes the **divisor function**

$$\sigma_s(n) = \sum_{0<d|n} d^s,$$

where the sum is over all positive divisors d of n. We can estimate $\sigma_s(n)$, using the Riemann zeta function, as follows:

$$\sigma_s(n) = \sum_{0<d|n} d^s = n^s \sum_{0<d|n} d^{-s} \leq n^s \zeta(s),$$

if $s > 1$.

The estimate for cusp forms comes from Deligne's proof of the Weil conjectures on zeta functions of n-dimensional projective nonsingular varieties over finite fields, as we mentioned earlier in our discussion of Ramanujan's tau function. Deligne's proof cannot be discussed here. Instead we give an exercise proving a weaker result. See Exercise 3.4.18. □

Exercise 3.4.18 (A Weak Estimate for the Fourier Coefficients of Cusp Forms). Show that if $f(z)$ is a cusp form of weight k for $SL(2,\mathbb{Z})$ as in Theorem 3.4.4, we have the inequality $|a_n| \leq Cn^{k/2}$, for some positive constant C.

Hint. *Since $f(z)$ vanishes at infinity, we can factor $\exp(2\pi iz)$ out of the Fourier expansion and obtain $|f(z)| \leq C\exp(-2\pi y)$. Since the function $g(z) = |f(z)|y^{k/2}$ is invariant under $SL(2,\mathbb{Z})$ and approaches zero as y approaches infinity, it is bounded by some constant M. It follows that the Fourier coefficient a_n satisfies $|a_n| \leq My^{-k/2}\exp(2\pi ny)$ for all $y > 0$. Take $y = 1/n$ to complete the proof.*

The idea of Exercise 3.4.18 is Hecke's. Classical improvements in the estimate were obtained by Rankin [534] and Selberg [570], among others. Note that the proof in Exercise 3.4.18 applies also to noncongruence subgroups and nonintegral weights, as well as to the nonholomorphic modular forms to be considered in the next section, while Deligne's method does not apply to these other situations.

There is an interesting question raised by Atkin and Swinnerton-Dyer. The Fourier coefficients for modular forms on congruence subgroups have "bounded denominators" in the sense that the forms are linear combinations of functions with coefficients that are integers in some algebraic number field (of finite degree over $\mathbb{Q}$). This is not the case for noncongruence subgroups.

Question. Do *all* of the modular forms for noncongruence subgroups have this "bad" property unless they are really modular forms on congruence subgroups?

3.4.9 The Representation of Integers by Quadratic Forms

Let P be a positive definite even integral matrix P of determinant one as in Theorem 3.4.3. Define the representation numbers $r_P(m)$ by

$$r_P(m) = \#\{\, a \in \mathbb{Z}^n \mid P[a] = 2m \,\}.$$

It follows from what we know about modular forms of weight k for $SL(2,\mathbb{Z})$ that there is a positive constant C such that for every $\varepsilon > 0$,

$$\left| r_P(m) - \frac{2k}{B_k}\sigma_{k-1}(m) \right| \leq Cm^{\frac{k-1}{2}+\varepsilon}, \text{ where } k = n/2,\ k \text{ even integral.} \tag{3.79}$$

Here B_k denotes the kth Bernoulli number defined in Sect. 1.4.

Exercise 3.4.19. Prove the estimate Eq. (3.79).

Hint. *From Theorem 3.4.3, you know that* $\theta(P,z) \in \mathcal{M}(SL(2,\mathbb{Z}),n/2)$. *So there is a constant* c_k *and a cusp form* f_k *such that*

$$\theta(P,z) = c_k G_k(z) + f_k(z).$$

Compare constant terms in the Fourier series for both sides of this equation to see that $c_k = [2\zeta(k)]^{-1}$. This allows you to evaluate the mth Fourier coefficient of $c_k G_k(z)$, using Exercise 3.4.2. Use Euler's formula for the Riemann zeta function at even integers, which was mentioned in Sect. 1.4 and Theorem 3.4.4 (and proved in Exercise 3.5.7) to finish the exercise.

It would be more interesting perhaps to obtain formulas for the representation numbers of $P = I =$ the $n \times n$ identity matrix. Then one would have a formula for the number of ways of representing m as a sum of n squares. Such results for certain n are considered in Rademacher [525, Chap. 11], and Knopp [357, Chap. 5]. Siegel answers much more general questions than these in his work on quadratic forms mentioned earlier.

Exercise 3.4.20. Show that if $n = 8$ and P is a positive definite 8×8 symmetric matrix satisfying the conditions of Theorem 3.4.3, then

$$r_P(m) = 240\sigma_3(m).$$

3.4.10 Korteweg–DeVries Equation

The nonlinear partial differential equation often written in the form $u_t + 6uu_x + u_{xxx} = 0$ was derived by Korteweg and DeVries in 1895 in a study of long water waves in a rectangular canal. If you replace u by $-u$, the equation becomes $u_t - 6uu_x + u_{xxx} = 0$. Often the constant 6 is replaced by 12 as well. Of course, in physical problems the constants would reflect the properties of the materials involved. The *KdV* equation applies to many other situations (e.g., plasma physics) in which solitons occur. This means that waves show a particle-like behavior in interactions; that is, the interacting waves keep their shape and amplitude while undergoing a shift. Korteweg and DeVries found a solution involving the hyperbolic secant and one involving elliptic functions. See the book of Lonngren and Scott

[429] for various articles on solitons and the *KdV* equation (particularly the articles of Miura and Hermann). Another reference is Kasman [339]. The mandatory quotation on the subject is from J. Scott-Russell's "Report on Waves" from 1844. I found it in Keener [348, p. 408].

> 'I was observing the motion of a boat which was rapidly drawn along a narrow channel by a pair of horses, when the boat suddenly stopped—not so the mass of water in the channel which had put it in motion; it accumulated round the prow of the vessel in a state of violent agitation, then suddenly leaving it behind, rolled forward with great velocity, assuming the form of a large solitary elevation, a rounded, smooth and well-defined heap of water, which continued its course along the channel apparently without change of form or diminution of speed. I followed on horseback, and overtook it still rolling on at a rate of some eight or nine miles an hour, preserving its original figure some thirty feet long and a foot to a foot and a half in height. Its height gradually diminished, and after a chase of one or two miles I lost it in the windings of the channel. Such, in the month of August 1834, was my first chance interview with that singular and beautiful phenomenon....

Exercise 3.4.21 (The Korteweg–DeVries Equation). Show that the Weierstrass $\wp$-function which satisfies the nonlinear differential equation given in formula (3.63) also represents a travelling wave solution of the Korteweg–DeVries equation $u_t - 6uu_x + u_{xxx} = 0$ of the form $u = w(x + ct)$. This gives after a couple of integrations the ODE: $(w')^2 = 2w^3 - cw^2 - gw - h$, where g and h are constants of integration. This ODE now begins to resemble the differential equation for the Weierstrass $\wp$-function. See Kasman [339] for more details.

3.5 Modular Forms: Not So Classical—Maass Waveforms

> Il a fallu Maass pour nous sortir du ghetto des fonctions holomorphes.[5]
>
> —From Weil [729, Vol. III, p. 463].

3.5.1 Maass Waveforms

Our main goal for the remainder of this chapter is to develop harmonic analysis on $SL(2,\mathbb{Z})\backslash H$. That is, we seek the spectral decomposition of the non-Euclidean Laplace operator Δ on H into $SL(2,\mathbb{Z})$-invariant eigenfunctions. This is a non-Euclidean analogue of Fourier series. However, because the quotient $SL(2,\mathbb{Z})\backslash H$ is not compact, there is a continuous spectrum as well as a discrete spectrum, so that one has a mixture of Fourier series and integrals. In this section, we shall study the eigenfunctions of Δ on $SL(2,\mathbb{Z})\backslash H$. We will find that these eigenfunctions have much in common with the classical automorphic or modular forms considered in the preceding section. So we shall also call these eigenfunctions of Δ "modular forms." But usually we will call them "Maass waveforms," because they were first systematically considered by Maass [439] in 1949. We will find that the only known *explicit* example of a Maass waveform for $SL(2,\mathbb{Z})$ is the Epstein zeta function from Sect. 1.4, which can be viewed as a nonholomorphic analogue of the Eisenstein series in Sect. 3.4. However, analogues of cusp forms are also known to exist, but we cannot produce examples such as the discriminant function Δ in Sect. 3.4, except for congruence subgroups (since the product of eigenfunctions of the Poincaré Laplacian (which we also call Δ) will not usually be an eigenfunction of Δ). However, computer approximations of Maass cusp forms are plentiful.

---

[5] We needed Maass to help us get out of the ghetto of holomorphic functions.

Some references for this section are Borel and Casselman [52], Borel and Mostow [54], Bump [71], Cartier and Hejhal [78], Delsarte [129, Tome II, pp. 599–601, 829–845], Elstrodt [155, 156], Faddeev [171], Fay [176], Gangolli [191], Gangolli and Warner [195, 196], Gelbart [200], Gelfand et al. [203], Goldfeld and Hundley [219], Goldfeld and Husemoller [220], Hejhal [261–263, 266], Huber [306, 307], Iwaniec [319], Kubota [375], Lachaud [382], Lang [389], Lax and Phillips [400], Maass [437, 439], Moreno [477], Roelcke [544–546], Sarnak [555, 556], Selberg [569–571, 573], Venkov [693–695], Weil [728], and Zagier [750].

Definition 3.5.1. A function $f: H \to \mathbb{C}$ is a **Maass waveform (or nonholomorphic modular or automorphic form)** if it satisfies the following three conditions:

(1) f is an eigenfunction of the non-Euclidean Laplacian; i.e., $\Delta f = \lambda f$, for some $\lambda \in \mathbb{C}$, where $\Delta = y^2(\partial^2/\partial x^2 + \partial^2/\partial y^2)$.

(2) f is invariant under the modular group; i.e., $f(\gamma z) = f(z)$ for all $\gamma \in \Gamma = SL(2,\mathbb{Z})$ and all $z \in H$.

(3) f has at most polynomial growth at infinity; i.e., there are constants $C > 0$ and k such that $|f(z)| \leq Cy^k$, as $y \to \infty$ uniformly in x.

Here $\mathcal{N}(SL(2,\mathbb{Z}),\lambda)$ will be our notation for the vector **space of such Maass waveforms**. Since $\Delta f = \lambda f$ is an elliptic partial differential equation, the Maass waveform f is automatically real analytic (see Garabedian [197, Chap. 5]). There is a discussion in Maass's lectures [437, Chap. 4] of a generalization of the Maass waveform which contains the classical holomorphic modular form from Sect. 3.3 as well.[6] A fundamental application of Maass waveforms for discrete subgroups Γ of $SL(2,\mathbb{R})$ with compact quotients $\Gamma\backslash H$ was already noted by Delsarte in 1942 [129, Tome II, pp. 599–601]. Then in the 1950s Selberg developed his trace formula, along with many number-theoretical applications, some of which were found independently by Eichler (see Selberg [569] and Eichler [152]). We shall discuss the trace formula in Sect. 3.7.

---

[6] Actually, you must multiply Maass's functions g in $[SL(2,\mathbb{Z}),s,s,1]$ by y^s in order to obtain our functions f in $\mathcal{N}(SL(2,\mathbb{Z}),s(s-1))$.

Example 3.5.1. **(Eisenstein Series—Alias Epstein's Zeta Function)** As in Sect. 3.2, the easiest way to construct eigenfunctions of Δ is to build them up out of the **power function**

$$p_s(z) = (\operatorname{Im} z)^s = y^s, \quad \text{if } z = x + iy \in H, \tag{3.80}$$

for clearly, $\Delta p_s = s(s-1)p_s$. Formulas (3.16) and (3.25), along with the exercises following those formulas, show that the K-Bessel and associated Legendre functions are built up out of power functions by integration over the appropriate subgroup of $SL(2,\mathbb{Z})$. In the present situation we must sum rather than integrate, since $SL(2,\mathbb{Z})$ is a discrete subgroup. The **Eisenstein series** $E_s(z)$ for $s \in \mathbb{C}$ with $\operatorname{Re} s > 1$ and $z \in H$, is defined by

$$E_s(z) = \sum_{\gamma \in \Gamma_\infty \backslash \Gamma} p_s(\gamma z), \tag{3.81}$$

where

$$\Gamma_\infty = \left\{ \begin{pmatrix} \pm 1 & * \\ 0 & \pm 1 \end{pmatrix} \in SL(2,\mathbb{Z}) = \Gamma \right\}. \tag{3.82}$$

The notation in Eq. (3.81) means that the sum runs over a complete set of representatives $\gamma \in \Gamma$ for the quotient $\Gamma_\infty \backslash \Gamma$. It is necessary to "mod out" Γ_∞, because

$$p_s(\gamma z) = p_s(z) \quad \text{for all} \quad \gamma \in \Gamma_\infty, z \in H.$$

Note that Γ_∞ is the **stabilizer** in Γ of the cusp at infinity (i.e., the subgroup of Γ fixing ∞). When there are more cusps in the fundamental domain for the discrete group Γ, there will be more Eisenstein series (one for each Γ-inequivalent cusp). The convergence in formula (3.81) is absolute and uniform on compacta in the s-plane, by arguments given in Sect. 1.4 for Epstein's zeta function, using the following exercise.

Exercise 3.5.1 (The Connection Between Eisenstein Series and Epstein Zeta Functions). Let Y be in

$$\mathcal{SP}_2 = \left\{ Y \in \mathbb{R}^{2\times 2} \;\middle|\; Y \text{ positive definite, } |Y| = 1 \right\}.$$

Define Epstein's zeta function for $\operatorname{Re} s > 1$, $Y \in \mathcal{SP}_2$, as in Sect. 1.4, by

$$Z(Y,s) = \frac{1}{2} \sum_{a \in \mathbb{Z}^2 - 0} Y[a]^{-s},$$

where $Y[a] = {}^t a Y a$, viewing $a \in \mathbb{Z}^2 - 0$ as a column vector.

Recall that, according to Exercise 3.1.9 of Sect. 3.1, we can identify H and $\mathcal{SP}_2$ via $z \mapsto W_z$ where $z = x + iy$ and $W_z \in \mathcal{SP}_2$ is defined by

$$W_z = \begin{pmatrix} 1/y & 0 \\ 0 & y \end{pmatrix} \begin{bmatrix} 1 & -x \\ 0 & 1 \end{bmatrix} = \begin{pmatrix} 1/y & -x/y \\ -x/y & (x^2+y^2)/y \end{pmatrix}.$$

(a) Show that if $a = \begin{pmatrix} n \\ -m \end{pmatrix} \in \mathbb{R}^2$, then

$$W_z[a] = {}^t a W_z a = y^{-1} |mz+n|^2.$$

(b) Show that if $\operatorname{Re} s > 1$, then the Eisenstein series (3.81) is given by

$$E_s(z) = \frac{y^s}{2} \sum_{(m,n)=1} |mz+n|^{-2s}.$$

Here $(m,n) =$ the greatest common divisor of m and n.

(c) Prove that $\zeta(2s)E_s(z) = Z(W_z, s)$, if $\zeta(s) =$ Riemann's zeta function.

(d) Show that $Z(W,s) = Z(W^{-1}, s)$ if $W \in \mathcal{SP}_2$.

Hints on part (b). *You must show that the quotient* $\Gamma_\infty \backslash \Gamma$ *has as representatives*

$$\begin{pmatrix} * & * \\ m & n \end{pmatrix}$$

with $(m,n) = 1$. *This is a consequence of the fact that*

$$\begin{pmatrix} \pm 1 & q \\ 0 & \pm 1 \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} * & * \\ \pm c & \pm d. \end{pmatrix}$$

You also need to know that a row vector in $\mathbb{Z}^2$ *can be completed to a matrix in* $SL(2,\mathbb{Z})$ *if and only if the* gcd *(=greatest common divisor) of its entries is one.*

Hints on part (c). *In the definition of* $Z(W,s)$, *the sum runs over all* $(m,n) \in \mathbb{Z}^2 - 0$. *Factor the greatest common divisor out of m and n.*

Hints on part (d). *If* W *is a* 2×2 *positive matrix of determinant* 1, *then*

$$W^{-1} \begin{bmatrix} c \\ d \end{bmatrix} = W \begin{bmatrix} -d \\ c \end{bmatrix}.$$

From Exercise 3.5.1, we see that $E_s \in \mathcal{N}(SL(2,\mathbb{Z}), s(s-1))$ when $\operatorname{Re} s > 1$, provided that $E_s(z)$ has polynomial growth as $y \to \infty$.

Exercise 3.5.2.

(a) Show that $E_s(\gamma z) = E_s(z)$ for all $\gamma \in SL(2,\mathbb{Z})$ and all $z \in H$, if $\mathrm{Re}\, s > 1$.
(b) Show that $E_s(z) \sim y^s$, as $y \to \infty$, for s fixed with $\mathrm{Re}\, s > 1$.
(c) Conclude that $E_s(z) \in \mathcal{N}(SL(2,\mathbb{Z}), s(s-1))$ when $\mathrm{Re}\, s > 1$.

Hint on part (b). *Use the formula from part (b) of Exercise 3.5.1.*

Theorem 1.4.1 of Sect. 1.4 shows that $E_s(z)$ has an analytic continuation to the entire complex s-plane as a meromorphic function of s. And $E_s(z)$ has a pole at $s = 1$ with residue $3/\pi = (\mathrm{vol}(SL(2,\mathbb{Z})\backslash H))^{-1}$. Moreover, the Eisenstein series satisfies the **functional equation**[7]

$$\Lambda(s)E_s(z) = \Lambda(1-s)E_{1-s}(z) \quad \text{if} \quad \Lambda(s) = \pi^{-s}\Gamma(s)\zeta(2s). \tag{3.83}$$

There must be a functional equation relating E_s and E_{1-s} because both functions lie in $\mathcal{N}(SL(2,\mathbb{Z}), s(s-1))$, which will be proved to be one-dimensional for $\mathrm{Re}\, s \neq \frac{1}{2}$.

Since $E_s(z) = Z(W_z, s)/\zeta(2s)$, any zero of $\zeta(2s)$ will produce a pole of $E_s(z)$ unless $Z(W_z, s)$ vanishes at that value of s. The trivial zeros of $Z(W_z, s)$ and $\zeta(2s)$ are both $s = -1, -2, -3, \ldots$; and both have order 1. Thus the trivial zeros cancel out in $E_s(z)$. The Euler product for the Riemann zeta function in formula (1.21) and Exercise 3.5.8 show that $\zeta(2s) \neq 0$ for $\mathrm{Re}\, s \geq \frac{1}{2}$. This is true in a larger region, but the Riemann hypothesis that the nontrivial zeros of $\zeta(2s)$ must lie on the line $\mathrm{Re}\, s = \frac{1}{4}$ remains unproved after more than 100 years. One of the motivations for the study of harmonic analysis on $\Gamma\backslash H$ is the desire to understand the complex zeros of $\zeta(s)$ (see Hejhal [262, p. 479] or [263], Cartier and Hejhal [78], or Zagier [750, p. 276]). Note that the zeros of $Z(W, s)$ depend on W and, in general, do not satisfy the Riemann hypothesis (see Stark [629] and Titchmarsh [680, p. 244]). They tend, however, to lie on the line $\mathrm{Re}\, s = \frac{1}{2}$, but can also be found in $\mathrm{Re}\, s > 1$.

From Exercise 3.5.2 and the functional equation (3.83) we see that

$$\Lambda(s)E_s(z) \sim \Lambda(s)y^s + \Lambda(1-s)y^{1-s} \quad \text{as } y \to \infty.$$

In fact, it is not hard to obtain the complete Fourier expansion of the Eisenstein series (see Exercises 3.5.3 and 3.5.4). The result of Exercise 3.5.4 is a very useful one in number theory for it reveals much about the Eisenstein series. It is thus reminiscent of the incomplete gamma expansion of Epstein's zeta function (Theorem 3.4.1 of Sect. 1.4). This might lead one to say that these formulas are the meat and potatoes (or quiche & salad, or tofu & rice) of the subject.

Exercise 3.5.3 (Fourier Expansions of Maass Waveforms). Show that a modular form $f \in \mathcal{N}(SL(2,\mathbb{Z}), s(s-1))$ has a Fourier series expansion as a periodic function of x given by

$$f(z) = ay^s + by^{1-s} + \sum_{n \neq 0} a_n \sqrt{y} K_{s-\frac{1}{2}}(2\pi|n|y) exp(2\pi i n x),$$

where $K_s(y)$ is the K-Bessel function in Sect. 3.2.

---

[7] We shall use the letter Γ for discrete groups such as $SL(2,\mathbb{Z})$ as well as for the gamma function. Hopefully the meaning will be clear from the context.

Hints. Use separation of variables in $\Delta f = s(s-1)f$, as in Exercise 3.5.4 of Sect. 3.2, and rule out the second solution $I_s(y)$ of the second-order ordinary differential equation for $K_s(y)$, by the polynomial growth of $f(z)$ as $y \to \infty$ (see Lebedev [401, p. 123]).

Exercise 3.5.4 (The Fourier Expansion of the Eisenstein Series). Set $\Lambda(s) = \pi^{-s}\Gamma(s)\zeta(2s)$ and $E_s^*(z) = \Lambda(s)E_s(z)$. Show that $E_s^*(z)$ has the Fourier expansion

$$E_s^*(z) = y^s\Lambda(s) + y^{1-s}\Lambda(1-s) + 2\sum_{n\neq 0} |n|^{s-1/2}\sigma_{1-2s}(n)y^{1/2}K_{s-1/2}(2\pi|n|y)e^{2\pi inx}.$$

Here $\sigma_s(n)$ is the **divisor function** defined by

$$\sigma_s(n) = \sum_{0<d|n} d^s,$$

where the sum runs over all positive divisors d of n.

Hints.
Method 1 (Chowla and Selberg [88]). Start with the Mellin transform result in Exercise 1.4.7 of Sect. 1.4:

$$E_s^*(z) = \frac{1}{2}\int_0^\infty t^{s-1} \sum_{(a,b)\in\mathbb{Z}^2-0} \exp\left(-\pi t W_z\begin{bmatrix} a \\ b \end{bmatrix}\right) dt. \tag{3.84}$$

Now, using the formula for W_z in Exercise 3.5.1, we have

$$W_z\begin{bmatrix} a \\ b \end{bmatrix} = y^{-1}(a-xb)^2 + yb^2.$$

Thus the $b=0$ term of the sum in Eq. (3.84) is integrated to obtain $\Lambda(s)y^s$. When $b \neq 0$, the variable a is summed over all of $\mathbb{Z}$, and thus one can apply the Poisson sum formula and obtain, as in Exercise 3.4.11 of Sect. 3.4

$$\sum_{a\in\mathbb{Z}} \exp\left[-\pi\frac{t}{y}(a-xb)^2\right] = \sqrt{\frac{y}{t}}\sum_{a\in\mathbb{Z}} \exp\left[-\pi\frac{y}{t}a^2 + 2\pi iabx\right] \tag{3.85}$$

After pulling out the $b=0$ term, substitute Eq. (3.85) into Eq. (3.84). Then use the integral formula for $K_s(z)$ in Exercise 3.2.1 of Sect. 3.2 to complete the proof.
Method 2 (Bateman and Grosswald [30] or Terras [665]). Apply the Poisson summation formula directly to the series defining $Z(W,s)$, after taking out the $b=0$ term of the sum over $(a,b)\in\mathbb{Z}^2-0$, as in Method 1.
Method 3 (Kubota [375, pp. 13–17]). Note that

$$E_s(z) = y^s + y^s \sum_{\substack{\gcd(c,d)=1 \\ c\geq 1}} |cz+d|^{-2s} = y^s + y^s\sum_{c\geq 1} c^{-2s} \sum_{\gcd(c,d)=1} |z+d/c|^{-2s}.$$

Let $d = r + cn,\ 0 \leq r < c,\ n \in \mathbb{Z}$. Use Poisson summation on the sum over n to complete the proof. You will also need the identity relating the divisor function and the singular series (see Hardy [249, p. 141]):

$$\zeta(2s) \sum_{\substack{c>0, d \bmod c \\ \gcd(d,c)=1}} c^{-2s} \exp(2\pi imd/c) = \sigma_{1-2s}(m).$$

The name "singular series" was used by Hardy and Littlewood in their work on Waring's problem (see Hardy [250, Vol. I, pp. 377–532]). It was applied to Fourier coefficients of Eisenstein series by Siegel [600, Vol. 1, p. 329] and Maass [438, pp. 300–313]. This third method is related to the Bruhat decomposition of $SL(2,\mathbb{Q})$ (see Volume II i.e., [667]).

The singular series form of the Fourier coefficients of $E_s(z)$ in Method 3 of the preceding exercise involves **Ramanujan sums**:

$$c_r(n) = \sum_{\substack{1 \leq m < r, \\ \gcd(m,r)=1}} \exp(2\pi imn/r).$$

Ramanujan studied various formal trigonometric series involving these sums (including that obtained from $E_s(z)$). Kac [330, pp. 86–96] showed that such series can often be viewed as Fourier expansions of almost periodic functions. See also Delsarte [129, Vol. II, pp. 603–624].

Exercise 3.5.5. Show that at $s = 1/2$ the poles of the constant term in the Fourier expansion of the Eisenstein series $E_s^*(z)$ in Exercise 3.5.4 cancel. Evaluate the resulting constant in terms of known constants like π, γ =Euler's constant, etc. That is, evaluate

$$\lim_{s \to \frac{1}{2}} \left(\Lambda(s) y^s + y^{1-s} \Lambda(1-s) \right), \quad \text{where} \quad \Lambda(s) = \pi^{-s} \Gamma(s) \zeta(2s).$$

Exercise 3.5.6 (Kronecker's Limit Formula). Use Exercise 3.5.4 to show that if W_z is as in Exercise 3.5.1

$$\lim_{s \to 1} \left\{ Z(W_z, s) - \frac{\pi}{2} \frac{1}{s-1} \right\} = \pi \left\{ \gamma - \log 2 - \log \left(\sqrt{y} \left| \eta\left(x+iy\right) \right|^2 \right) \right\}$$

where $\gamma =$ **Euler's constant** defined by

$$\gamma = \lim_{n \to \infty} \left(\sum_{m=1}^{n} \frac{1}{m} - \log n \right) = 0.577215\ldots,$$

and $\eta(z) =$ Dedekind's eta function from Sect. 3.4.

Hint. *Note that* $K_{\frac{1}{2}}(y) = (\pi/2y)^{\frac{1}{2}} e^{-y}$ *and use the definition of eta as an infinite product.*

The result of Exercise 3.5.6 goes back to Kronecker [374, Vol. 4, pp. 222, 347–495, Vol. 5, pp. 1–132], along with a second limit formula for the function

$$\sum_{a\in\mathbb{Z}^2-0} W[a]^{-s}\exp(2\pi i\,{}^t qa) \quad \text{for } W \in \mathcal{SP}_2,\ q \in \mathbb{R}^2.$$

Interesting discussions of this result can be found in Siegel [599], Weber [724, Vol. 3], and Weil [727]. The Kronecker limit formulas are of central importance for the construction of class fields of algebraic number fields, as Kronecker had already demonstrated in 1863, by proving that the limit formula gives a solution $(x,y) \in \mathbb{Z}^2$ of Pell's equation

$$x^2 - dy^2 = \pm 1 \quad (\text{if } d \text{ is a given positive integer})$$

in terms of modular functions eta and theta. Some related references are Goldstein [224], Hecke [258, pp. 198–207, 290–312], Katayama [340], Lang [390], Meyer [466], Ramachandra [529], Shintani [594], and Zagier [751]. Stark [626, part 4] uses Kronecker's limit formula to prove Stark's conjectures on values of L-functions when the base field is $\mathbb{Q}$ or an imaginary quadratic field. This gives explicit reciprocity laws in some abelian extensions of the base field. Stark [630] and Gupta [237] use the Kronecker limit formula to obtain an analytic proof of the Coates–Wiles theorem on L-functions for elliptic curves with complex multiplication.

There are other sorts of applications of the Fourier expansion in Exercise 3.5.4. Stark [625] uses generalizations of that result to show that there are exactly nine imaginary quadratic fields with class number 1. The Fourier expansion can also be used to obtain comparisons of values of the Riemann zeta function at even and odd positive integers $2n$ and $2n+1$ in terms of rapidly converging series of exponentials (see Terras [666] and Exercise 3.5.7). One has Euler's formula for $\zeta(2n)$, $n = 1,2,3,\ldots$, but there is no simple interpretation for $\zeta(2n+1)$, $n = 1,2,3,\ \ldots$. There are, however, conjectures of Lichtenbaum relating these odd values to orders of K-groups (see Lichtenbaum [420]). In Volume II [667] we will find that $\zeta(3)$ does appear in the volume of the fundamental domain for $SL(3,\mathbb{Z})$. This fact was already noted by Siegel, but the Gauss–Bonnet theorem does not apply in this case to relate $\zeta(3)$ to a rational number times some power of π (see Weil [729, Vol. I, p. 561]). Apéry showed that $\zeta(3)$ is indeed irrational (see Van der Poorten [690]).

Exercise 3.5.7.

(a) Prove **Euler's formula**:

$$\zeta(2n) = \frac{(-1)^{n-1}}{2(2n)!}(2\pi)^{2n}B_{2n}, \qquad n = 1,2,3,\ldots.$$

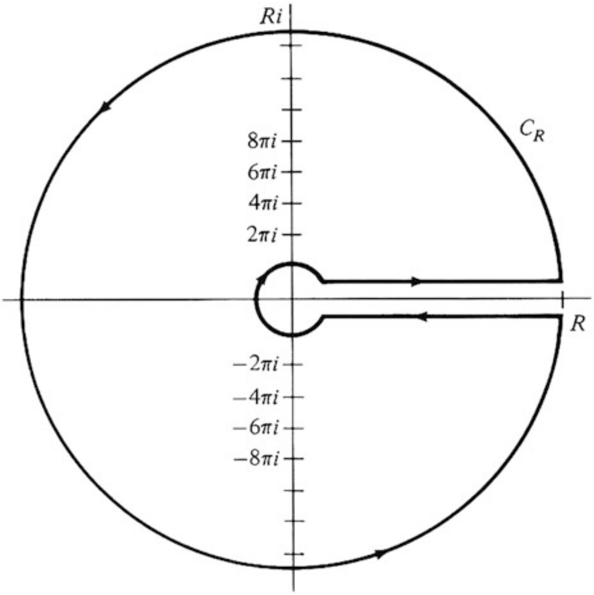


Fig. 3.37 Contour for the evaluation of $\zeta(2n)$, $n \in \mathbb{Z}^+$

Here B_n= the **nth Bernoulli number** defined by

$$\frac{x}{e^x - 1} = \sum_{n \geq 0} \frac{B_n}{n!} x^n.$$

(b) Use Exercise 3.5.4 to show that

$$\zeta(3) = \frac{2}{45}\pi^3 - 4\sum_{n \geq 1} e^{-2\pi n}\sigma_{-3}(n)\left(2\pi^2 n^2 + \pi n + \frac{1}{2}\right).$$

(c) Show that

$$\lim_{s \to 1}\left(\zeta(s) - \frac{1}{s-1}\right) = \gamma = \text{ Euler's constant.}$$

Hint on Part (a). *Look at the contour C_R in Fig. 3.37 and let $R \to \infty$ in*

$$\int_{C_R} (e^z - 1)^{-1} z^{-n} dz,$$

using the Cauchy integral formula and residue calculus.
Hint on Part (b). *See Terras [666].*

Exercise 3.5.8. Show that $\zeta(1+it) \neq 0$ for all t in $\mathbb{R}$.

Hints. (See Jacquet and Shalika [324] and Zagier [750].) Suppose that $\zeta(1+it) = 0$. Then $E_{(1+it)/2}(z)$ is of rapid decay as $y \to \infty$, according to Exercise 3.5.4. But then if $s = (1+it)/2$, we have

$$\begin{aligned} C(E_s, y) &= \int_{x=0}^{1} E_s(x+iy)\, dx \\ &= (\text{the constant term in the Fourier expansion of } E_s) = 0. \end{aligned}$$

Consider

$$\begin{aligned} I(E_s, r) &= \int_{y=0}^{\infty} C(E_s, y) y^{r-2} dy \\ &= \int_{\Gamma_\infty \backslash H} E_s(z) y^{r-2} dx\, dy = \int_{\Gamma / H} E_s(z) E_r(z) y^{-2}\, dx\, dy. \end{aligned}$$

Here Γ_∞ is as in Eq. (3.82). Let $r = \overline{s}$. Then $I(E_s, \overline{s})$ is the square of the L^2-norm of E_s and we have a contradiction.

The result in Exercise 3.5.8 is quite old and is fundamental for proofs of the prime number theorem using Tauberian theorems (see Wiener [736, pp. 112–121]). As Selberg [571] noticed, the preceding exercise is easy once one has the analytic continuation of $E_s(z)$ to Re $s \geq \frac{1}{2}$ with the only pole occurring at $s = 1$, for one need only examine a nonconstant Fourier coefficient of $E_s(s)$ [which has $\zeta(2s)$ in the denominator] to see that $\zeta(2s)$ cannot vanish on Re $s = \frac{1}{2}$. The usual proofs of Exercise 3.5.8 are quite different (see Grosswald [233, p. 131]).

Related Fourier expansions of Eisenstein series have been used by Kubota [376] as well as by Heath-Brown and Patterson [257] to study cubic Gauss sums. One can prove Gauss's conjecture that the average order of the number of classes of positive integral binary quadratic forms of discriminant $-D$ is $2\pi D^{1/2}/(7\zeta(3))$ by interpreting class numbers as Fourier coefficients of holomorphic Eisenstein series of weight $\frac{3}{2}$ (see Hecke [258, pp. 499–504]). Goldfeld et al. [218] use similar ideas to study the arithmetic of elliptic curves.

Exercise 3.5.9.

(a) Let $BC^\infty(\Gamma\backslash H)$ denote the set of bounded C^∞ functions on $\Gamma\backslash H$. Show that if Δ denotes the Poincaré Laplacian and f, Δf both lie in $BC^\infty(\Gamma\backslash H)$ then (assuming that f is real-valued) we have:

$$(\Delta f, f) = -\int\limits_{\Gamma\backslash H} \left(\left(\frac{\partial f}{\partial x}\right)^2 + \left(\frac{\partial f}{\partial y}\right)^2 \right) dx\, dy \leq 0,$$

where the inner product is

$$(f,g) = \int\limits_{\Gamma\backslash H} f(z)\overline{g(z)}y^{-2}dxdy.$$

(b) Deduce that if λ is an eigenvalue of Δ on $L^2(\Gamma\backslash H)$, $\lambda = s(s-1)$, then $\operatorname{Re} s = \frac{1}{2}$ or $s \in [0,1]$.

(c) Suppose that $f,g \in BC^\infty(\Gamma\backslash H)$ are real-valued functions such that Δf and Δg are also in $BC^\infty(\Gamma\backslash H)$. Then Δ is symmetric; i.e.,

$$(\Delta f, g) = (g, \Delta f).$$

Hint. *(See Lang [389, pp. 281–284]) Use Green's theorem on a truncated fundamental domain as in Fig. 3.38.*

Using the notation of Exercise 3.5.9, let D_Δ be the space of functions f in $BC^\infty(\Gamma\backslash H)$ such that Δf is also in $BC^\infty(\Gamma\backslash H)$. Then D_Δ is a dense subspace of $L^2(\Gamma\backslash H)$ and Δ can be extended to a self-adjoint operator (see Lang [389, pp. 284–287]). This is done using the resolvent $(\Delta - s(s-1))^{-1}$. Exercise 3.5.9 shows that the operator is negative.

3.5.2 Maass Cusp Forms

There are (sometimes) other members of $\mathcal{N}(\Gamma,\lambda)$ besides Eisenstein series. Recalling what happened in the last section for holomorphic modular forms, we expect to see cusp forms. We will have existence theorems and computer approximations but sadly so far none of the many explicit examples like Δ from the last section. The only explicit examples of Maass cusp forms come from Hecke L-functions with grossencharacters and they are for congruence subgroups not the full modular group (see Maass [439] and Hejhal and Strömbergsson [275] who call them CM Maass forms).

Definition 3.5.2. We say that $f \in \mathcal{N}(\Gamma,\lambda)$ is a **Maass cusp form** for $\Gamma = SL(2,\mathbb{Z})$ if the constant term in the Fourier expansion in Exercise 3.5.3 vanishes; i.e.,

$$\int\limits_{x=0}^{1} f(x+iy)dx = 0 = ay^s + by^{1-s} \qquad \text{for all } y > 0.$$

Let $\mathcal{SN}(\boldsymbol{\Gamma},\boldsymbol{\lambda})$ denote the **vector space of Maass cusp forms** in $\mathcal{N}(\Gamma,\lambda)$.

Exercise 3.5.10. Show that if f is a cusp form then $f \in L^2(\Gamma\backslash H)$, using the invariant area element $y^{-2}dx\,dy$.

Hint. *From the Fourier series for f and the asymptotic behavior of $K_s(y)$ as $y \to \infty$, you can show that a cusp form must be bounded and thus is square-integrable since the fundamental domain has finite area (see Sect. 3.3). The Fourier series for f converges uniformly on the fundamental domain by the Weierstrass M-test, once you have a bound on the Fourier coefficients (see Theorem 3.5.2).*

The Eisenstein series E_s is not a cusp form.

Exercise 3.5.11. Show that if $\Lambda(W,s) = \pi^{-s}\Gamma(s)Z(W,s)$, for $W \in \mathcal{SP}_2$, $s \neq 0,1$, then

$$\begin{aligned} &\Lambda(W,s) \in L^1(\Gamma\backslash H, y^{-2}dxdy) \quad \text{if} \quad 0 < \operatorname{Re} s < 1 \quad \text{(and not otherwise);} \\ &\Lambda(W,s) \notin L^2(\Gamma\backslash H, y^{-2}dxdy) \qquad \text{for all } s. \end{aligned}$$

Hint. *As in Exercise 3.5.10, make use of the Fourier series for $\Lambda(W,s)$ in Exercise 3.5.4.*

Now we want to study the structure of the space of Maass forms $\mathcal{N}(SL(2,\mathbb{Z}),\lambda)$ more closely, recalling what happened for classical modular forms in Sect. 3.4. Given f in $\mathcal{M}(SL(2,\mathbb{Z}),k) =$ the classical holomorphic modular forms of weight k, there is always a constant c such that $f - cG_k$ is a cusp form, where $G_k =$ the holomorphic Eisenstein series. This is not so clear in the nonholomorphic case, since the Fourier expansion of f in $\mathcal{N}(SL(2,\mathbb{Z}),s(s-1))$ begins with $ay^s + by^{1-s}$, rather than with a constant. Suppose that Re $s > \frac{1}{2}$, $s \notin [0,1]$. Then we can find a constant c such that $(f - cE_s)$ has constant term by^{1-s}. This implies that $(f - cE_s)$ is square-integrable over the fundamental domain, using the non-Euclidean invariant area element $y^{-2}dx\,dy$. But this contradicts the fact that the Laplace operator is negative on the fundamental domain (see Exercise 3.5.9). We have thus proved part (a) of the following theorem.

Theorem 3.5.1.

(a) If $\operatorname{Re} s > \frac{1}{2}$, *and* $s \notin [\frac{1}{2},1]$, *then*

$$\mathcal{N}(SL(2,\mathbb{Z}),s(s-1)) = \mathbb{C}E_s.$$

(b) $\mathcal{N}(SL(2,\mathbb{Z}),0) = \mathbb{C}$.
(c) If $\operatorname{Re} s = \frac{1}{2}$ *or* $s \in [\frac{1}{2},1)$, *then*

$$\mathcal{N}(SL(2,\mathbb{Z}),s(s-1)) = \mathbb{C}E_s \oplus \mathcal{SN}(SL(2,\mathbb{Z}),s(s-1)).$$

Note. Later we shall prove that if $s \in [\frac{1}{2},1)$, *then* $\mathcal{SN}(SL(2,\mathbb{Z}),s(s-1)) = \{0\}$ *(see Theorem 3.5.3).*

Proof. These results would be easy if the constant term in the Fourier expansion of a Maass waveform did not have two parts, for you could easily subtract a multiple of the Eisenstein series to produce a cusp form. And a harmonic cusp form (i.e., a

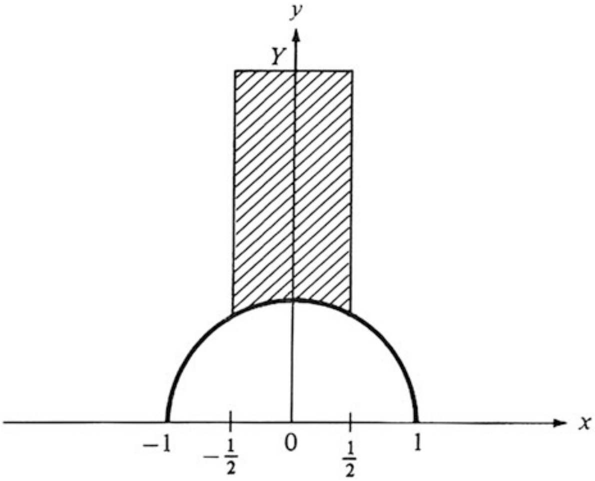


Fig. 3.38 Fundamental domain truncated at $y = Y$

cusp form f such that $\Delta f = 0$) would have to be identically zero by the maximum principle for harmonic functions (see Garabedian [197]).

In order to circumvent this difficulty, we must prove the one-dimensionality of the vector space V consisting of constant terms in the Fourier expansion of $f \in \mathcal{N}(SL(2,\mathbb{Z}),\lambda)$. One proof of this fact uses Green's theorem on the truncated fundamental domain

$$D_Y = \left\{ z \in H \;\middle|\; |z| \geq 1,\ \operatorname{Re} z \in \left[-\frac{1}{2}, \frac{1}{2}\right], \operatorname{Im} z \leq Y \right\}$$

pictured in Fig. 3.38. Suppose that $f \in \mathcal{N}(SL(2,\mathbb{Z}),\lambda)$ and $g \in \mathcal{N}(SL(2,\mathbb{Z}),\mu)$ with Fourier expansions

$$f(z) = \sum_{m\in\mathbb{Z}} a_m(y)\exp(2\pi imx),\ g(z) = \sum_{m\in\mathbb{Z}} b_m(y)\exp(2\pi imx).$$

Then the Euclidean version of Green's theorem on the region D_Y says that if $\partial/\partial n$ is the normal derivative on the boundary ∂D_Y, and ds the differential of arc length, we have

$$\begin{aligned}
&\textstyle\int_{D_Y} (f\Delta g - g\Delta f)\frac{1}{y^2}dx\,dy = \int_{\partial D_Y}\left(f\frac{\partial g}{\partial n} - g\frac{\partial f}{\partial n}\right)ds \\
&= \int_{x=-1/2}^{+1/2} \sum_{n,m\in\mathbb{Z}} \left(a_n(Y)b_m'(Y)e^{2\pi i(n+m)x} - a_n'(Y)b_m(Y)e^{2\pi i(n+m)x}\right)dx \\
&= \sum_{m\in\mathbb{Z}} \left(a_m(Y)b_{-m}'(Y) - a_m'(Y)b_{-m}(Y)\right).
\end{aligned}$$

Here we have used the fact that f and g are invariant under $SL(2,\mathbb{Z})$ to see that the integral over the boundary ∂D_Y reduces to the integral over the top horizontal line $\operatorname{Im} z = Y$, since the $SL(2,\mathbb{Z})$ identifications make the top the only real boundary of the manifold D_Y. The last formula comes from the orthogonality of the exponentials $\{\exp(2\pi inx),\, n \in \mathbb{Z}\}$.

The explicit expressions for the Fourier coefficients of f and g from Exercise 3.5.3 imply that if $m \neq 0$ and $\lambda = \mu = s(s-1)$, then

$$0 = a_m b'_{-m} - a'_m b_{-m}.$$

Therefore, $\lambda = \mu = s(s-1)$ implies that $0 = a_0 b'_0 - a'_0 b_0$. Moreover, we know that

$$a_0(y) = ay^s + cy^{1-s} \quad \text{and} \quad b_0(y) = by^s + dy^{1-s}.$$

When $s \neq \frac{1}{2}$, it follows that $ad - bc = 0$, which means that the dimension of the vector space V of constant terms is indeed 1. The dimension is obviously 1 in the case that $s = \frac{1}{2}$. □

Note. You can think of the constants in part (b) of Theorem 3.5.1 as residues of the Eisenstein series at $s = 1$.

The following exercise will be useful in the last section of this chapter when we evaluate the parabolic term of the Selberg trace formula. It gives an idea of the independence of the two Eisenstein series as distributions.

Exercise 3.5.12.

(a) Show that if $\varphi(s) = \Lambda(1-s)/\Lambda(s)$, where $\Lambda(s) = \pi^{-s}\Gamma(s)\zeta(2s)$, then

$$\begin{aligned}\int_{D_A} E_s E_{s'} \frac{1}{y^2}\,dxdy &= \frac{(s'-s)A^{s+s'-1}}{(s-s')(1-s'-s)} - \frac{\varphi(s)\varphi(s')A^{1-s-s'}(s'-s)}{(s-s')(1-s'-s)} \\ &\quad + \frac{A^{s-s'}\varphi(s')(1-s-s')}{(s-s')(1-s'-s)} + \frac{A^{s'-s}\varphi(s)(s'+s-1)}{(s-s')(1-s'-s)} + o(1) \\ &= \left\{ \frac{A^{s+s'-1}}{s+s'-1} - \frac{A^{1-s-s'}\varphi(s)\varphi(s')}{s+s'-1} \right\} \\ &\quad + \frac{A^{s-s'}\varphi(s')}{s-s'} - \frac{A^{s'-s}\varphi(s)}{s-s'} + o(1), \quad as \quad A \to \infty.\end{aligned}$$

Here D_A is the truncated fundamental domain in Fig. 3.38 and E_s denotes the Eisenstein series, as usual. The notation "little-o of 1", $o(1)$, stands for a function which approaches 0, as $A \to \infty$.

(b) Let $s' = \overline{s}$, $s = \sigma + ir$, in part (a) and take the limits as $\sigma \to \frac{1}{2}$ to show that the term in braces from part (a) approaches

$$\frac{1}{2}\left\{4\log A - \frac{\varphi'}{\varphi}\left(\frac{1}{2}+ir\right) - \frac{\varphi'}{\varphi}\left(\frac{1}{2}-ir\right)\right\}$$

Hint. *See Kubota [375]. Start with* $\int\limits_{D_A}((\Delta E_s)E_{s'} - E_s(\Delta E_{s'}))\frac{1}{y^2}dxdy$. *Use Green's theorem.*

Theorem 3.5.1 is a special case of a result of Maass [439, pp. 169–170] (see also Maass [437, pp. 195–215]). Formulas similar to those in Exercise 3.5.12 have been called the **Maass–Selberg relations** by various authors (see Kubota [375, pp. 18–20], Harish-Chandra [252, p. 75], Hejhal [261, Vol. II], Langlands [393, p. 333]). However, the Maass–Selberg relations involve truncating the Eisenstein series rather than the fundamental domain. See the website of Paul Garrett for many pages on the subject:

http://www.math.umn.edu/~garrett/

He says: "The (false) assumption that truncated Eisenstein series are essentially eigenfunctions does permit a plausible looking and traditional (!) derivation of the correct inner product of two truncated Eisenstein series… but, of course, also of any other assertion, true or false. This outcome can be and has been (erroneously) construed as validating the (incorrect) argument that truncated Eisenstein series are effectively eigenfunctions."

The Maass–Selberg relations will be used in Sect. 3.7 to evaluate the parabolic terms in Selberg's trace formula. These relations are also used in many of the references mentioned above to obtain the analytic continuation of the Eisenstein series for very general discrete groups, a result that was quite easy for $SL(2,\mathbb{Z})$ (see Theorem 1.4.1 of Sect. 1.4). It is also possible to give inner-product formulas for truncated Eisenstein series and then no $o(1)$ term appears (see Selberg [571, pp. 183–184]). Selberg [574, Vol. I, p. 673] contains some interesting commentary on the Maass–Selberg relations, which draws a distinction between the result of Maass and that of Selberg.

3.5.3 Computations of Maass Cusp Forms

Now, what can be said of the spaces of Maass cusp forms for $SL(2,\mathbb{Z})$? We know from Exercise 3.5.9 and Theorem 3.5.1 that this space of cusp forms is $\{0\}$ unless s lies in the interval $[0,1]$ or the real part of s is $\frac{1}{2}$. Furthermore, Theorem 3.5.3 will show that, in fact, the space is also $\{0\}$ when $s \in (0,1)$. We shall show in Sect. 3.7 that $\mathcal{SN}(SL(2,\mathbb{Z}), s(s-1)) \neq \{0\}$ for an infinite number of values s. However, no one has ever produced an exact value of s for which the space of cusp forms is nonzero—much less an example of a cusp form for the modular group.

Nevertheless, it *is* possible to obtain computer approximations of cusp forms. Note that a cusp form vanishes like a constant times $\exp(-2\pi y)$ as y approaches infinity. Thus, it is not too far wrong to consider the problem of finding eigenfunctions of the non-Euclidean Laplacian on the compact region D_Y in Fig. 3.38. The boundary conditions would be the periodicity coming from $f(z+1) = f(z)$ and $f(-1/z) = f(z)$, plus the vanishing of $f(z)$ on the horizontal line $y = Y$. Note, further, that the symmetry $u(x+iy) = -x+iy$ leaves the domain D_Y and the Laplacian invariant. Thus the symmetry splits the space of solutions to our boundary value problem into even and odd functions of $\mathrm{Re}\, z$. The space of odd functions will satisfy the Dirichlet problem, requiring them to vanish on the boundary. The space of even functions will satisfy the Neumann problem requiring that their normal derivatives vanish on the boundary.

Exercise 3.5.13. Prove the last statements about even and odd eigenfunctions of Δ on the fundamental domain for $SL(2,\mathbb{Z})$.

Computations of the eigenvalues of the non-Euclidean Laplacian for the Dirichlet and Neumann problems have been done since at least the 1970s. The earliest computations were made by Cartier [77], Cartier and Hejhal [78], Haas [244], Hejhal [263, 266, 273], Hejhal and Berg using a CRAY-I supercomputer (see Hejhal [261, Appendix C to Vol. II]), Hejhal and Rackner [274], and Stark [631]. The most recent computations were made by Strömberg [644], Booker et al. [48], and Then [677]. The last reference finds Maass cusp form for eigenvalues bigger than 1.6×10^9. Now you can find lists of eigenvalues and Fourier coefficients of the corresponding eigenforms on various websites such as LMFDB, the database of L-functions, modular forms, and related objects

http://www.lmfdb.org/ModularForm

or

http://modform.org/ModularForms/MaassForms.

There are also density plots on various websites; e.g., those of Hejhal, Strömberg, and Strömbergsson.

We give some older eigenvalue tables here. The first table is that of Hejhal and Berg. See Table 3.5. This table has a shorter list of eigenvalues for cusp forms corresponding to the Neumann problem because the numerical method that was used produces "spurious eigenvalues" for the Neumann problem. These spurious eigenvalues would fill in the blanks for Neumann eigenvalues, except that they correspond to spurious cusp forms with a logarithmic singularity at the point $\exp(2\pi i/3)$ in the fundamental domain. Hejhal [263] proves this and shows that the s corresponding to the spurious eigenvalue $\lambda = s(s-1)$ are exactly the zeros of the Dedekind zeta function of the algebraic number field $\mathbb{Q}(\exp(2\pi i/3))$.

See Sect. 1.4 for a discussion of the Dedekind zeta function. The particular one that arises in Hejhal [263] is actually also an Epstein zeta function as well as the product of $\zeta(s)$ and $L(s,(-3/*))$, formed with the Kronecker symbol $(-3/*)$. See the discussion after formula (3.74) of Sect. 3.3 for references on the Kronecker symbol. The L-function appearing here is defined by

Table 3.5 Hejhal–Berg table of eigenvalues λ of Δ on $\mathcal{H}/SL(2,\mathbb{Z})$, for $\lambda = s(s-1),\ s=\frac{1}{2}+it$

t for odd cusp forms		t for even cusp forms
Dirichlet problem		Neuman problem
9.53369 52613 536		13.77975 13518 907
12.17300 83246 797		17.73856 33811
14.35850 95182 59		19.42348 147
16.13807 31715 23		21.31579 6
16.64425 92018 8		
18.18091 78346		
19.48471 38		
20.10669 4		
21.47905 7		
22.19467		
24.41971	incomplete	
25.05085	below	
26.0568	double line	
26.4469		
27.28		

The last digit in each number is uncertain. From Hejhal [261, Vol. II, pp. 653, 730]

Table 3.6 Zeros $s=\frac{1}{2}\pm i\gamma$ of $\zeta(s)L(s,(-3/*))$

γ for $\zeta(s)$	γ for $L(s,(-3/*))$
14.134725	8.039737
21.022040	11.249206
25.010858	15.704619
30.424876	18.261997
32.935062	20.455771

$$L(s,(-3/^*)) = \sum_{n\geq 1}(-3/n)n^{-s} \quad \text{for} \quad \operatorname{Re} s > 1.$$

This is a very special example of a class of L-functions associated to Dirichlet characters. For the analytic continuation and functional equations of such L-functions, see for example Davenport [120]. The general theory is developed in Lang [388]. One can find tables of zeros of Dirichlet L-functions (see Spira [613]) and of Riemann's zeta function (see Haselgrove and Miller [256] and Brent [65]). Or you can look at the L-functions websites mentioned above. We list a few of these zeros in Table 3.6.

Table 3.5 should be compared with Table 3.8 which was made by Cartier [77] and lists eigenvalues for the non-Euclidean Laplacian on a rectangle. It is also interesting to compare Table 3.5 with Table 3.7 of Haas. The main difference between the Haas table and that of Hejhal and Berg is the absence in the latter table of the Neumann eigenvalues corresponding to the zeros of the Dedekind zeta function just mentioned. The spurious eigenvalues in the Haas table come from the numerical method used by Haas as well as Hejhal and Berg. All start with the Fourier expansion of an even cusp form:

Table 3.7 The Haas table of eigenvalues λ of Δ on $\mathcal{H}/SL(2,\mathbb{Z})$, $\lambda = s(s-1)$, $s = \frac{1}{2} + it$ as corrected by Hejhal to contain 18.261997 (which had been omitted by mistake)

t for Dirichlet's problem	t for Neumann's problem
9.533695	8.039738
12.17301	11.24921
14.35851	13.77975
16.13807	14.13473
16.64426	15.70462
18.18092	17.73856
19.48471	18.261997
20.10669	19.42348
	20.45578

This table includes spurious eigenvalues in the Neumann list as is explained in the text. The spurious eigenvalues correspond to numbers in Table 3.6

$$u(z) = \sum_{n>0} c_n y^{1/2} K_{it}(2\pi n y)\cos(2\pi n x)$$

and seek to find the unknown Fourier coefficients c_n and the number t, by choosing N point $z_1, \ldots, z_N$ in H and solving

$$\frac{u(z_j) - u(z_j^*)}{|z_j - z_j^*|} = 0, \quad j = 1, \ldots, N, \quad z^* = 1/\bar{z}.$$

The condition can be rewritten as a system of N linear equations in the N unknowns c_i, provided that one forgets the terms after c_N in the Fourier expansion of the cusp form u. Thus t must satisfy

$$\det(I_i(z_j, t))_{1 \le i,j \le N} = 0 \quad \text{for } I_j(z,t) \text{ involving } K\text{-Bessel functions.}$$

Hejhal proves in the above-mentioned paper that the eigenvalues which appear on the Haas list and not on the Hejhal list (the zeros of the Dedekind zeta function of $\mathbb{Q}(\exp(2\pi i/3))$ are spurious because they correspond to eigenfunctions that behave exactly like cusp forms except that they have a logarithmic singularity at the point $\exp(2\pi i/3)$ in the fundamental domain. So the Fourier expansion that is produced by this numerical method will converge only for $y > \sqrt{3}/2$ and not for $y \le \sqrt{3}/2$. One can draw a very interesting moral on the care required in using computers (especially when dealing with Neumann boundary value problems with continuous spectra).

Table 3.8 Cartier's table of eigenvalues $\lambda = s(s-1)$, $s = \frac{1}{2} + it$, for Δ on the rectangle $R = \{x + iy \,|\, 0 < x < \frac{1}{2}, 1 < y < 5\}$

t for Dirichlet's problem	t for Neumann's problem
9.790	8.906
12.421	11.631
14.590	13.863
16.522	15.837
17.019	16.108
18.291	17.644
19.948	19.330
20.309	19.465

Hejhal also notes that the Fourier coefficients c_n in the expansion

$$f(z) = \sum_{n \neq 0} c_n \exp(2\pi inx) y^{1/2} K_{it}(2\pi|n|y)$$

for a spurious eigenform $f(z)$ with $c_1 = 1$ do not satisfy the inequality

$$|c_n| \leq \sigma_0(n) n^{3/10}, \tag{3.86}$$

where $\sigma_0(n)$ is the number of positive divisors of n. However, as we shall see in Theorem 3.6.4 of Sect. 3.6, we can assume that the cusp form $f(z)$ is an eigenfunction for all the Hecke operators and we can normalize $f(z)$ so that the first Fourier coefficient $c_1 = 1$. For such forms, the inequality (3.86) has been proved (see Moreno [477] and Selberg [570]). This gives another indication that the spurious cusp forms are suspect.

The inequality (3.86) is a weak form of the **Ramanujan–Petersson conjecture for cuspidal Maass waveforms which are normalized eigenfunctions of the Hecke operators** (to be defined in Sect. 3.6). This conjecture says

$$\left|c_p\right| \leq 2 \quad \text{for all primes } p. \tag{3.87}$$

Such a result is not contained in Deligne's theorem stated in Theorem 3.4.4 of Sect. 3.4. Better bounds than that given in Eq. (3.86) have been obtained, but Eq. (3.87) is still unproved as we write this (see Marie-France Vignéras [703] for a discussion of the improvement of the exponent $\frac{3}{10}$ in Eq. (3.86) to $\frac{1}{4}$, as well as a discussion of the connections with group representations). It is now known that we can replace 3/10 by 7/64. See Kim [351]. Tables 3.9 and 3.10 provide a small amount of numerical evidence for the Ramanujan–Petersson conjecture (3.87), by listing the first few Fourier coefficients corresponding to various Maass waveforms. Check the extensive tables on the website of Strömbergsson for more evidence.

There are now lots of tables of Fourier coefficients of Maass cusp forms available on the web. Booker et al. [48] say: "We have computed the first ten eigenvalues on $PSL(2,\mathbb{Z})\backslash H$, namely,

Table 3.9 The Hejhal–Berg table of Fourier coefficients of cusp forms for the eigenvalue λ of Δ on $H/SL(2,\mathbb{Z})$, $\lambda = s(s-1), s = \frac{1}{2} + it$ (c_n denoting the nth Fourier coefficient)

b for even cusp forms	c_2	c_3	c_5	c_7
13.77975 13518 907	1.54930 44779 4	0.24689 97724 5	0.73706 04	−0.2614
17.73856 33811	−0.76545 80566	−0.97777 89075	−1.01527 35	1.1807
19.42348 147	−0.69276 198	1.56235 43	−0.03841 2	0.313
21.31579 6	1.28752 9	1.25177 3	1.170	−0.54
t for odd cusp forms	c_2	c_3	c_5	c_7
9.53369 52613 536	−1.06833 35512	−0.456197 354	−0.29067 256	−0.7449
12.17300 83246 797	0.28925 18714	−1.20185 8761	0.03955 272	0.4481
14.35850 95182 59	−0.23091 51912	0.69559 49863	−1.29828 45	−0.4834
16.13807 31715 23	1.16185 5592	−1.28197 2561	−0.75680 63	−0.2985
16.64425 92018 8	−1.54022 7825	0.97749 2591	−0.10524 2	−0.693
18.18091 78346	0.37406 3346	0.10195 8698	0.63733 1	−1.542
19.48471 38	−1.70018 80	−0.61456 54	0.8198	0.063
20.10669 4	0.85884 4	0.18727 7	−1.395	0.78
21.47905 7	−0.65625 1	0.22644 2	1.802	0.42
22.19467	1.59685	−1.11648	−0.637	−1.00

The last digit in each number is doubtful. From Hejhal [261, Vol. II, pp. 653, 730]

$$r \approx 9.5337, 12.1730, 13.7798, 14.3585, 16.1381,$$
$$16.6443, 17.7386, 18.1809, 19.4235, 19.4847,$$

to a precision of more than 1,000 decimal digits, together with the first 455 Fourier coefficients $a_1, \ldots, a_{455}$ to 900 digits (at least the first 50 of these were actually obtained to more than 1,000 digits).” They also note that “The computer time required was between one and three weeks per example … (on a 1.5 GHz PC).” Here we still give the old tables of Hejhal and Berg (see Table 3.9) and Stark (see Table 3.10). Hejhal used a supercomputer while Stark used the UCSD VAX.

The numerical methods used by Hejhal and most of the others are limited by the decreasing size of the K-Bessel functions in the Fourier expansions of cusp forms. Accurate evaluations of the K-Bessel functions are essential. Riho Terras [675] did this for the papers I wrote involving K-Bessel expansions in the 1970s (see Purdy et al. [523]). Now the K-Bessel function is available in most computer packages such as Mathematica. However, those who do the massive computation of Maass forms still have to worry. Booker et al. [48] say “the most time-consuming task is, by far, that of computing the values of the K-Bessel function.”

Hejhal appends this comment of C. L. Siegel to his table:

> Ausserdem ist in Erwägung zu ziehen, dass die oben angegebenen numerischen Werte von $v(x)$ mit Hilfe von Rechenmaschinen bestimmt wurden und daher ebenfalls im strengen Sinne unbewiesen sind." —From Siegel [600, Vol. III, p. 439]).[8]

In producing Table 3.10 of Fourier coefficients for the first even nonholomorphic cusp form, Stark [631] used the fact that the cusp form is an eigenfunction for the Hecke operators to be considered in the next section. Thus we have the equation

$$c_p f(z) = p^{-\frac{1}{2}} \sum_{j \bmod p} f\left(\frac{z+j}{p}\right) + p^{-\frac{1}{2}} f(pz)$$

(see Theorem 3.6.4 of Sect. 3.6). Stark then fixes a point such as $z \cong 1.4i$ and looks at the **Hecke points**

$$z_j = (z+j)/p,\ j = 1, \ldots, p. \tag{3.88}$$

One can find a matrix $A_j \in SL(2,\mathbb{Z})$ such that $w_j = A_j z_j$ lies in fundamental domain for $SL(2,\mathbb{Z})$. So our equation becomes

$$c_p f(z) = p^{-\frac{1}{2}} \sum_{j \bmod p} f(w_j) + p^{-\frac{1}{2}} f(pz). \tag{3.89}$$

One then uses Eq. (3.89) to obtain the coefficients c_p recursively with more and more accuracy. You calculate $f(z)$ and the other values of f using the Fourier expansion of f, with a few Fourier coefficients; e.g., $c_1 = 1$. Then you get some digits of c_2. You plug those in and get more digits. This assumes that you know the eigenvalue, but you can also use the process to approximate eigenvalues by checking that the function f approximated is invariant under the substitution $z \to -1/z$.

Figure 3.39 shows a plot of the points w_j used by Stark in formula (3.89) above for $p = 983$ after they have been moved (using the map $z \to -1/z$) from the standard fundamental domain with its cusp at ∞ to the fundamental domain with a cusp at 0. One can then ask whether these points $w_j, j \bmod p$, become uniformly distributed (with respect to the non-Euclidean area) in the fundamental domain as p approaches infinity. This would account for the fact that the sum (3.89) of p terms times $p^{-\frac{1}{2}}$ seems to be giving a result that is smaller by a factor of $p^{-\frac{1}{2}}$ than one would expect. In fact, the Ramanujan–Petersson conjecture for Maass waveforms would follow from such uniform distribution of the points w_j in the fundamental domain. Indeed, Stark has suggested that perhaps one could use the Ramanujan–Petersson conjecture for holomorphic modular forms to show that the desired cancellation by $p^{-\frac{1}{2}}$ does occur. See Chiu [86] where it is proved that the Hecke points become dense in the fundamental domains of $SL(2,\mathbb{Z})$ and $SL(3,\mathbb{Z})$ as $p \to \infty$. See also Exercise 3.6.17.

---

[8] One also has to keep in mind that the numerical values of $v(x)$ were determined with the help of a calculator and have consequently not been proved in a strict sense.

Table 3.10 Stark's table of Fourier coefficients for the first even cusp form

p	$c(p)$	p	$c(p)$	p	$c(p)$
2	1.5493044779	269	−1.4815318734	617	−0.9413471724
3	0.2468997725	271	0.5040655357	619	−1.2022515538
5	0.7370603853	277	0.0799184237	631	1.2204182342
7	−0.2614200758	281	1.2661555488	641	−1.6267782096
11	−0.9535646526	283	−1.6137265420	643	−1.2717803924
13	0.2788270292	293	−0.9962645200	647	1.3158148270
17	1.3073417145	307	1.1244326539	653	−0.7308513965
19	0.0925585825	311	−0.0039967298	659	−1.2242624630
23	1.1380685214	313	0.4789645104	661	−0.0234437218
29	0.7521138455	317	−1.5835570331	673	1.3778855357
31	0.0248519535	331	1.1152062270	677	0.0265995820
37	0.1992656556	337	0.4390031691	683	0.5292669101
41	−0.3040329968	347	−1.4315675896	691	−1.3160485151
43	0.7832393635	349	−0.3501353058	701	1.5888219871
47	0.3605684105	353	1.7625605824	709	−0.5251706918
53	1.3980657196	359	0.5325184967	719	−1.0362134271
59	−1.5877309619	367	−0.0033416821	727	−0.7502921970
61	1.1672759688	373	−0.9876462546	733	−0.9208587111
67	−0.0270938863	379	1.5933050198	739	0.8958405935
71	−1.6334238582	383	0.0502016847	743	1.6323865759
73	1.0678477369	389	−0.3273589916	751	−0.9959260442
79	−0.5311738889	397	−1.8931512743	757	1.3259008900
83	−0.9045323799	401	−0.8126034389	761	−0.5856192114
89	−1.0673531716	409	1.4753959508	769	−0.1033879313
97	−0.0032571155	419	0.7396796406	773	−1.4675416297
101	0.8641852969	421	0.8792514066	787	−0.4231228583
103	−1.2318448417	431	−1.2083667833	797	0.1724453075
107	−0.8126520455	433	−0.1571375484	809	−0.3987763947
109	−0.1537416810	439	−0.6020990403	811	−0.9581311790
113	0.8117725443	443	1.0132298238	821	−0.5238603585
127	1.1696610310	449	0.4461683165	823	0.1600228277
131	−0.6111258748	457	−0.7081760488	827	−0.1779829984
137	0.7718026731	461	0.3300809302	829	0.4892801545
139	0.0953561766	463	1.2726492305	839	1.5943055357
149	−0.1869675229	467	0.1481146844	853	−0.0944388002
151	−0.2836096280	479	1.1482791938	857	1.2009724917
157	1.1464542083	487	−1.1381929615	859	−0.9219874648
163	1.5036129830	491	−1.4887418682	863	1.1751819412
167	−0.7243411282	499	0.4359170070	877	−0.4805639483
173	0.0377358321	503	0.3463320737	881	−1.1622088814
179	−0.5924790734	509	0.7903047596	883	1.4610158527
181	1.8928188557	521	−0.8278509888	887	−1.9462665557
191	0.4549235996	523	−1.0508097534	907	−0.1627440731
193	−0.3510653631	541	0.6476703225	911	0.5225227593

(continued)

Table 3.10 (continued)

197	−0.4934900086	547	0.0132913465	919	−1.0727170339
199	0.7615593277	557	−0.8512529528	929	−0.1305562742
211	1.6334955452	563	0.7157534392	937	−1.0585435062
223	−0.8898352549	569	−0.6677981282	941	−0.7542322227
227	−1.1170228371	571	−0.2379643425	947	1.3539019798
229	−1.1231294363	577	0.9500705634	953	0.4262552906
233	1.1318288732	587	1.0559415688	967	0.7553483561
239	−0.6112502497	593	0.1197484790	971	−0.0605613598
241	1.6168818500	599	0.8007081710	977	−0.3377517502
251	−0.4397565620	601	1.2754657449	983	−1.7315964075
257	0.3781399469	607	−1.5299526484	991	1.6842084188
263	1.0274020100	613	0.6464422791	997	−1.0060566454

From Stark [631]. Reproduced with permission from Ellis Horwood, Chichester, England. Normalized eigenvalue ≈13.7797513519.

Figure 3.40 is a representative density plot of a Maass cusp form from an older version of Dennis Hejhal's website:

http://www.math.umn.edu/~hejhal/.

while Fig. 3.41 shows some density plots of Maass wave forms from Sarnak [560].

Physicists have asked whether distributions of eigenfunctions of Δ would "become localized" as the eigenvalue approaches infinity. By localizing, we mean that they tend to vanish on some open set in the fundamental domain. This idea came from looking at various sorts of billiard tables in the plane. If instead we look at density plots of Maass waveforms such as those in Figs. 3.40 or 3.41 from Sarnak, we will tend to think that the distributions do not localize. Well, I have to admit that I sometimes see monsters in these pictures and sometimes approximations to parts of horizontal lines or circles. But the horocycles and geodesics can bounce around quite madly. Thus even if the support of the function were concentrated on horocycles or geodesics, I still would not see an open set where the function vanishes in the limit of large eigenvalues.

To get a clearer idea what is happening, let's just take our discussion from Sarnak [560]. Suppose M is a Riemannian manifold with finite volume, negative curvature, and without boundary. Denote the Laplacian of M by Δ and the corresponding volume element by $d\mu$ normalized to have $\mu(M) = 1$. Suppose $\Delta\varphi + \lambda\varphi = 0$ on M, $\varphi \in L^2(\Omega)$. In quantum mechanics $d\nu_\varphi = |\varphi|^2 d\mu$ is interpreted as the probability distribution associated with being in the state φ. Do these measures become equidistributed as $\lambda \to \infty$ or can they localize? Sarnak notes: "If M has (strictly) negative curvature, then the geodesic flow is very well understood thanks to works of Hopf, Morse, Sinai, Bowen, and others. It is ergodic and strongly chaotic in all senses. The periodic geodesics are isolated and are unstable, and there is no restriction on how they may distribute themselves as their period increases...." See the last section for more information on geodesics in the fundamental domain.

In his work with Rudnick [548] and others, Sarnak was led to the **Quantum Unique Ergodicity conjecture (or QUE conjecture)**:

$$\nu_{\varphi} \to \mu \quad \text{as} \quad \lambda \to \infty.$$

Sarnak [560] says: "Put another way, μ is the only quantum limit." In [325, 422, 548, 612] it is proved that the QUE conjecture holds for arithmetic manifolds such as $SL(2,\mathbb{Z})\backslash H$ both for continuous and discrete series Maass waveforms. An analogous result for holomorphic modular forms (which are Hecke eigenfunctions) has been obtained by Holowinsky and Soundararajan [299] as the weight goes to infinity.

Some History

In the spring of 1977, Riho Terras (referred to as *mon ex-mari* in Cartier and Hejhal [78]) and I spent a personally disastrous but academically exciting sabbatical at the University of Bonn. During that time we visited Paris and R.T. spoke to Cartier about computing K-Bessel functions to improve the eigenvalue programs. We were lucky enough to witness a general strike and R.T.'s talk had to be cancelled.

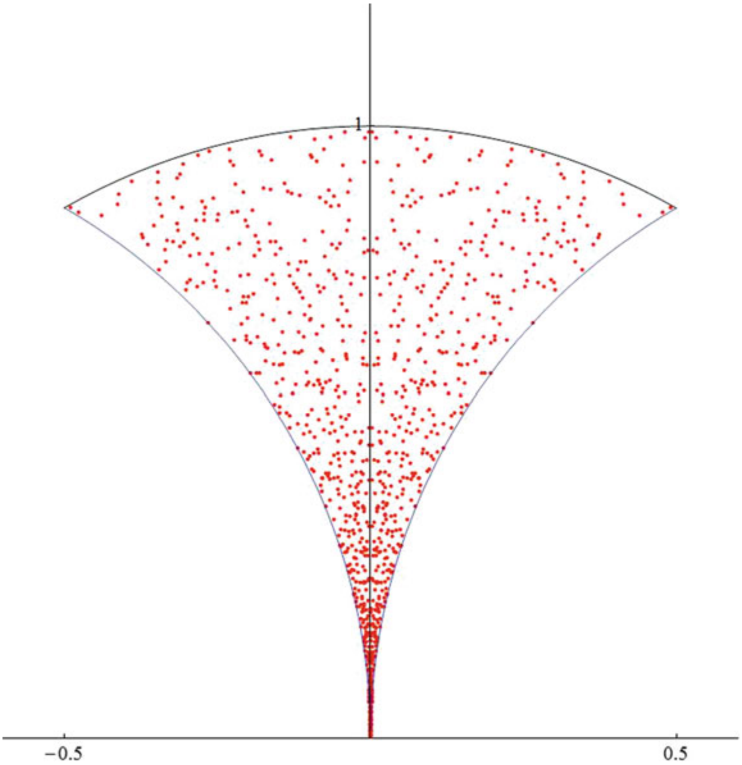


Fig. 3.39 Images of Hecke points from formula (3.88) in a fundamental domain for the modular group obtained by applying the map $z \to -1/z$ to the standard fundamental domain. The points were used by Stark in the computation of Table 3.10. Here we take $p = 983$ and $z = 2i$

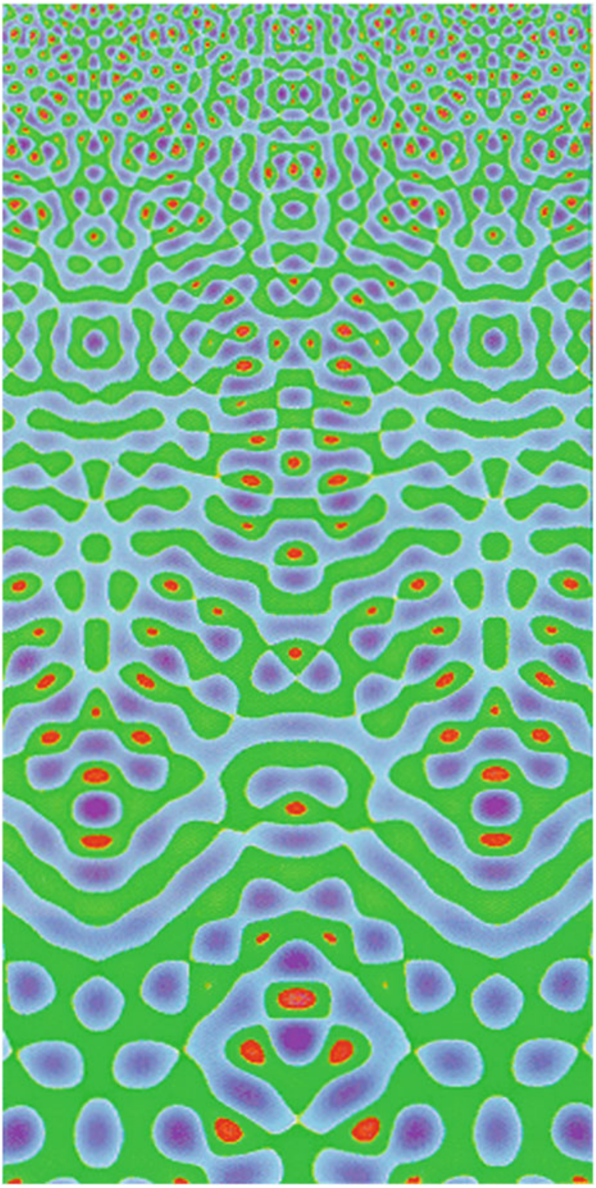


Fig. 3.40 A density plot of a Maass cusp form from an older version of Dennis Hejhal's website

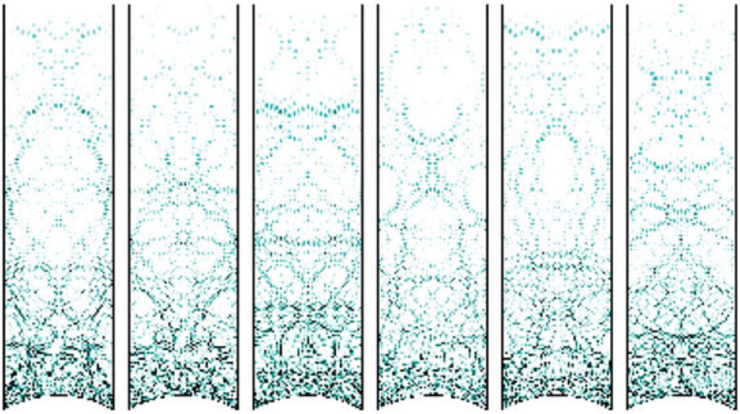


Fig. 3.41 Density plots of some Maass wave forms from Sarnak [560]

Later that spring I visited H. Maass in Heidelberg and met H. Neuenhöffer, who showed me the calculations of Haas. No one mentioned $\zeta(s)$ and my thoughts were only on the eigenvalue problem (and my own troubles). Then Neuenhöffer attempted to send me the Haas manuscript, but only the outside envelope arrived. Was someone in the Post Office interested in the Riemann hypothesis? Unknown! But finally a letter from Neuenhöffer reached me which included the Haas tables.

I put some of these tables in the first version of this book, on which I lectured at M.I.T. in the Fall of 1978. Then H. Stark noticed the zeros of $\zeta(s)$ in the Haas table. But he overlooked the L-function zeros somehow. By that time probably 100 people had a preliminary version of this book with the Haas table in it. But very few people looked at it, perhaps. Anyway, ultimately D. Hejhal visited U.C.S.D. and I showed him the table. He immediately said that it gave him a headache. Evidently he went to the library and found the L-function zeros also, but he did not tell me that. Instead he wrote a very cryptic remark on my blackboard, which he does not want printed here. But I think that he did not believe the Haas table even then. He worked hard for several months, and the result was Hejhal [263], which was announced in the summer of 1979 at the Durham conference.

In the preliminary version of this book I had attempted to use the Courant minimax principle to compare eigenvalue problems such as those represented by Tables 3.7 and 3.8 (see Courant and Hilbert [111, p. 409]). However, the existence of a continuous spectrum makes these arguments go awry.

One might still wonder whether there is any chance to prove the Riemann hypothesis by interpreting the zeros $s = \frac{1}{2} + it$ of $\zeta(s)$ as giving eigenvalues $s(s-1)$ of some self-adjoint operator. Pólya and Hilbert evidently suggested this independently around 1915, though they were probably not thinking of the eigenvalue problem discussed here. Hejhal's work appears to have laid to rest the possibility that Δ is the correct operator unless one can somehow deal with spurious eigenfunctions with logarithmic singularities. The fact that Epstein zeta functions do not satisfy the Riemann hypothesis in most cases leads one to become dubious about this approach. In fact, the Epstein zeta function for the matrix $\begin{pmatrix} 1 & 0 \\ 0 & 5 \end{pmatrix}$ has an infinite number of zeros in the region Re $s > 1$ (see Titchmarsh [680, p. 244]). See Venkov [695, p. 159] for a published version of a conjectural meaning for the eigenvalues associated to Maass cusp forms and their analogues. More recently there have been adelic versions of the Pólya–Hilbert conjecture by Paul Cohen and Alain Connes [102–104].

Exercise 3.5.14. Try to check Table 3.5.

Note. Hejhal says the preceding Exercise is too hard. A class of beginning physics and engineering graduate students agreed when they tried the standard finite-element method on the problem. You need good K_{it} programs if you use the Fourier expansion method that Hejhal used. It would also be interesting to program Stark's method using Hecke operators from Sect. 3.6 to produce Table 3.10.

Maass [439] shows that there are cusp forms for congruence subgroups of $SL(2,\mathbb{Z})$ [but not for $SL(2,\mathbb{Z})$ itself] which arise from Hecke L-functions of real quadratic number fields. See also Hejhal and Strömbergsson [275] who call them CM Maass forms, where CM stands for complex multiplication, referring to a phenomenon that happens for elliptic curves. Marie-France Vignéras [699] derives this in another way by integrating a certain theta function. We will return to this topic in the next section. It does not appear that anyone has been able to use such constructions to obtain cusp forms for $SL(2,\mathbb{Z})$, however.

Hejhal [273] computes Maass wave forms for Hecke triangle groups and finds no even cusp forms for the nonarithmetic Hecke triangle groups.

3.5.4 *Elementary Estimates of Fourier Coefficients and Eigenvalues*

We have mentioned the Ramanujan–Petersson conjecture for Maass cusp forms (which are eigenfunctions for the Hecke operators of the next section). This conjecture was stated in formula (3.87). Let us now prove some easier estimates on Fourier coefficients.

Theorem 3.5.2 (Estimates for Fourier Coefficients of Maass Waveforms).

(a) Let $E_s^*(z) = \pi^{-s}\Gamma(s)\zeta(2s)E_s(z)$ *and*

$$E_s^*(z) = ay^s + by^{1-s} + \sum_{n\neq 0} a_n\sqrt{y}K_{s-\frac{1}{2}}(2\pi|n|y)\exp(2\pi inx).$$

Then, for every $\varepsilon > 0$ *there is a positive constant* C_ε *such that*

$$|a_n| \leq C_\varepsilon |n|^{\mathrm{Re}\, s-\frac{1}{2}+\varepsilon}.$$

(b) *Suppose that the Maass cusp form* $f \in \mathcal{SN}(SL(2,\mathbb{Z}), s(s-1))$ *has the Fourier expansion*

$$f(z) = \sum_{n \neq 0} a_n \sqrt{y} K_{s-\frac{1}{2}}(2\pi|n|y)\exp(2\pi inx).$$

Then there is a positive constant C such that

$$|a_n| < C|n|^{\frac{1}{2}}.$$

Proof.

(a) We know from Exercise 3.5.4 that, if $\mathrm{Re}\, s > 1$, then

$$|a_n| = 2|n|^{\mathrm{Re}s-\frac{1}{2}}|\sigma_{1-2s}(n)| \leq 2|n|^{\mathrm{Re}s-\frac{1}{2}}\zeta(2\ \mathrm{Re}\ s-1).$$

However, if $\frac{1}{2} \leq \mathrm{Re}s \leq 1$, then $|a_n| < 2|n|^{\mathrm{Re}s-\frac{1}{2}}\sigma_0(n)$. Then Hardy and Wright tell us that given $\varepsilon > 0$, there is a positive constant C_ε such that $\sigma_0(n) < C_\varepsilon n^\varepsilon$ (see Hardy and Wright [251, pp. 260–262]).

(b) The proof proceeds as in Exercise 3.4.18 of Sect. 3.4. Use

$$a_n y^{1/2} K_{s-\frac{1}{2}}(2\pi|n|y) = \int_0^1 f(x+iy)\exp(-2\pi inx)dx,$$

and the fact that cusp forms are bounded on the upper half-plane. Then one sets $y = c/|n|$, where c is chosen so that $K_{s-\frac{1}{2}}(2\pi c) \neq 0$, to complete the proof.

□

Exercise 3.5.15. Fill in the details in the proof of Theorem 3.5.2.

Note that in Theorem 3.5.2, when $\mathrm{Re}\, s = \frac{1}{2}$, one has a better estimate for the Fourier coefficients of Eisenstein series than one has for cusp forms. This is rather strange if we remember the holomorphic case. There are better estimates than that given in Theorem 3.5.2 when the cusp form is an eigenfunction for all the Hecke operators. Then the Ramanujan–Petersson conjecture (3.87) above would say that the same sort of estimate would then hold. See the earlier discussion of Eqs. (3.86) and (3.87) for some references on Fourier coefficients of cusp forms. In addition, one can consult Bruggeman [68–70], and Kuznetsov [381].

Theorem 3.5.3. *Suppose* $\mathcal{SN}(SL(2,\mathbb{Z}),\lambda) \neq \{0\}$, *then* $\lambda \leq -3\pi^2/2$.

Proof (Roelcke [545, Sect. 7]). Suppose that

$$u(z) = \sum_{n \neq 0} a_n(y)\exp(2\pi inx) \in \mathcal{SN}(SL(2,\mathbb{Z}),\lambda).$$

Let D denote the usual fundamental domain for $SL(2,\mathbb{Z})$, given in Fig. 3.14. And let $D^* = \gamma(D)$, where $\gamma(z) = -1/z$. Then if $(u,u) = 1$, as in Exercise 3.5.9, part (a), we have

$$\begin{aligned}
2|\lambda| &= -2(\Delta u, u) = \int_{D\cup D^*} \left(\left|\frac{\partial u}{\partial x}\right|^2 + \left|\frac{\partial u}{\partial y}\right|^2 \right) dx\, dy \\
&\geq \int\limits_{\substack{|x|\leq \frac{1}{2} \\ y \geq \sqrt{3/2}}} |u_x|^2 dx\, dy = \int\limits_{y\geq\sqrt{3}/2} \sum_{n\neq 0} 4\pi^2 n^2 |a_n(y)|^2 dy \\
&\geq 3\pi^2 \int\limits_{y\geq\sqrt{3}/2} \sum_{n\neq 0} |a_n(y)|^2 \frac{1}{y^2} dy = 3\pi^2 \int\limits_{y\geq\sqrt{3}/2} \int\limits_{|x|\leq\frac{1}{2}} |u(z)|^2 \frac{1}{y^2} dx\, dy \\
&\geq 3\pi^2 \int_D |u|^2 \frac{1}{y^2} dx\, dy = 3\pi^2.
\end{aligned}$$

This completes the proof. □

Note that Hejhal's Table 3.5 gives a much better estimate for the first eigenvalue corresponding to a cusp form. However, Theorem 3.5.3 is of interest, because the argument can be applied to a few congruence subgroups of small level.

Let $|\lambda_1(\Gamma)|$ denote the smallest nonzero eigenvalue of the Laplacian in absolute value on $\Gamma\backslash H$ such that the corresponding eigenfunction is orthogonal to the constants. You might wonder whether there are discrete groups Γ in $SL(2,\mathbb{R})$ such that $0 < |\lambda_1(\Gamma)| \leq \frac{1}{4}$. Selberg [570] gives examples of such groups Γ. Randol [530] proves that such groups Γ exist with $\Gamma\backslash H$ compact and having as many eigenvalues λ in $(-a,0)$ as you wish, for any given a. Moreover, Selberg [570, pp. 13–14] conjectured that if Γ is a congruence subgroup, then $|\lambda_1(\Gamma)| \geq \frac{1}{4}$. This is now called **Selberg's eigenvalue conjecture**. Selberg actually proved that $|\lambda_1(\Gamma)| \geq 3/16 \cong 0.1875$. Sarnak [556] says "I think this is the fundamental unsolved analytic question in modular forms. It has many applications to classical number theory .... If true, it is sharp. In the first place, the continuous spectrum ... begins at 1/4." See also Luo et al. [436] who show that $|\lambda_1(\Gamma)| \geq 171/784 \cong 0.2181$.

Elstrodt [156], Hejhal [261, Vol. II], and Marie-France Vignéras [703] give more history of this problem and many more references. In particular, Vignéras notes that this conjecture and that of Ramanujan and Petersson given in Eq. (3.87) above are "deux volets d'une même conjecture" [9] in representation theory (see also Satake's article in Borel and Mostow [54, pp. 261–262]). Piatetskii-Shapiro [516] notes that the statement in adelic representation theoretic language can fail, but not for $GL(n)$, using results of Jacquet and Shalika.

---

[9] Two parts of the same conjecture.

3.5.5 Dimensions of Spaces of Maass Cusp Forms

There is a **conjecture** that $\dim_{\mathbb{C}} \mathcal{SN}(SL(2,\mathbb{Z}),\lambda) \leq 1$ (see Cartier [77] as well as Booker et al. [48, p. 24]). All the eigenvalue tables created so far give evidence in favor of this suspicion. See Randol [531] for a proof that for sufficiently large eigenvalues, the dimensions of spaces of cusp forms for congruence groups $\Gamma(p)$, p an odd prime, must be greater than 1, using knowledge of the degrees of irreducible representations of the finite simple group $PSL(2,\mathbb{Z}/p\mathbb{Z})$. Here we prove only that the $\mathcal{SN}(SL(2,\mathbb{Z}),\lambda)$ are finite-dimensional vector spaces.

Theorem 3.5.4. *The vector space of Maass cusp forms* $\mathcal{SN}(SL(2,\mathbb{Z}),\lambda)$ *is finite dimensional.*

Proof (Maass [439, pp. 154–156], using an idea of Siegel [600, Vol. II, pp. 97–137]). Let the Maass cusp form f be in $\mathcal{SN}(SL(2,\mathbb{Z}),s(s-1))$. Suppose that the Fourier coefficients in Exercise 3.5.3 are all zero, up to the coefficients a_n with $|n| = m$. Then we can show that if m is sufficiently large (depending on t, when $s = \frac{1}{2} + it$), it forces f to be identically zero. We know that f is bounded on the upper half-plane by Exercise 3.5.10. Thus $|f(z)|$ has a maximum at some point $z_0 \in H$. Set $z_0 = x_0 + iy_0$ and note that

$$a_n y_0^{1/2} K_{it}(2\pi|n|y_0) = \int\limits_{x=-\frac{1}{2}}^{\frac{1}{2}} e^{2\pi inx} f\left(x + \frac{iy_0}{2}\right) dx \left(\frac{\sqrt{2}K_{it}(2\pi|n|y_0)}{K_{it}(\pi|n|y_0)}\right).$$

It follows that

$$M = |f(z_0)| < 2\sqrt{2}M \sum_{|n|>m} K_{it}(2\pi\,|n|\,y_0)/K_{it}(\pi\,|n|\,y_0).$$

The asymptotic formula for the K-Bessel function (see Exercise 3.2.2 of Sect. 3.2 or Lebedev [401, p. 123]) shows that if m is sufficiently large (depending on t and y_0), the preceding series of quotients of K-Bessel functions can be estimated by exponentials. The result is that

$$M \leq CM \exp(-\pi m y_0).$$

If m is larger than $(\log C)/\pi y_0$, it follows that M is zero and, thus, that f is identically zero, as was to be proved. □

Exercise 3.5.16. Use the error term in the asymptotic expansion for $K_{it}(y)$ to obtain a bound for the dimension of the space $\mathcal{SN}(SL(2,\mathbb{Z}), -(t^2 + \frac{1}{4}))$. This bound will be an increasing function of t.

We can obtain an asymptotic formula for the sums of dimensions of spaces of Maass cusp forms with eigenvalues less than or equal to x, as x approaches infinity, once we have proved the Selberg trace formula (see Sect. 3.7). The trace formula implies that if $N(x)$ denotes the number of eigenvalues λ_n of Δ corresponding to cusp forms, counted with multiplicity, such that $|\lambda_n| \le x$, then

$$N(x) \sim \frac{\text{area}\,(SL(2,\mathbb{Z})\backslash H)}{4\pi} x = \frac{x}{12} \quad \text{as} \quad x \to \infty. \tag{3.90}$$

The asymptotic formula (3.90) is the non-Euclidean analogue of the Weyl asymptotic law for the distribution of eigenvalues of the Euclidean Dirichlet problem in a compact domain in $\mathbb{R}^n$ (see Theorem 1.3.5 of Sect. 1.3 and Theorem 3.7.5 of Sect. 3.7).

In particular, formula (3.90) implies that there are infinitely many Maass cusp forms for $SL(2,\mathbb{Z})$. This application of the trace formula was pointed out by Selberg [573]. The argument only works for discrete subgroups Γ of $SL(2,\mathbb{Z})$ such that $\Gamma\backslash H$ is noncompact of finite volume, with the property that there are good bounds on the constant term coefficients in the Fourier expansion of the Eisenstein series on the line $\operatorname{Re} s = \frac{1}{2}$. Roelcke [545] gives more examples of groups Γ with infinitely many cusp forms. So there arose a conjecture, called the **Roelcke–Selberg conjecture**, to the effect that there are infinitely many cusp forms for very general discrete groups. Venkov has proved the infinite dimensionality of spaces of cusp forms with more general transformation properties for a wide class of discrete groups (see Venkov [693,695]). But Phillips and Sarnak [514] cast doubts on the conjecture for nonarithmetic groups. Hejhal's computations [273] for Hecke triangle groups seem to add some evidence in that Hejhal found no even cusp forms for nonarithmetic Hecke triangle groups. There are always odd cusp forms for these groups by analogous arguments to those in Exercise 3.7.6.

It is natural to ask whether there are connections between nonanalytic and analytic cusp forms. We will find that there are even more analogies in the next section. There are certainly various constructions that allow one to go from forms of one type to those of another type. Often this is done using a theta function for indefinite quadratic forms as a kernel and integrating (see Kudla [377, 378], and Marie-France Vignéras [699]). Theta functions of indefinite quadratic forms and related matters are also discussed in Siegel [600, Vol. III, in papers nos. 55, 58, 60]. A differential operator that maps nonholomorphic automorphic forms satisfying higher-order differential equations into holomorphic forms (in certain cases) is given by Schwandt [564]. See also Roelcke [544, Part I, Sect. 6] and Neuenhöffer [494]. Neuenhöffer gives a construction of nonholomorphic cusp forms via residues of Poincaré series.

Jacquet and Langlands gave an adelic version of a 1-1 correspondence between cusp forms for quaternion groups and cusp forms for congruence subgroups. Hejhal [272] gave a classical version of this correspondence for cuspidal Maass waveforms by integrating against a theta function for indefinite quadratic forms. See also Strömbergsson [645].

3.5.6 *Finite Analogues of Eisenstein Series*

We have discussed finite upper half-planes H_q over the finite field $\mathbb{F}_q$. Throughout this subsection **we assume that q is odd**, as we did in the Sect. 3.3.8 on finite upper half-planes. There we considered fundamental domains and tessellations of H_q for subgroups Γ of $GL(2,\mathbb{F}_q)$. Thus it is natural to consider analogues of modular forms for such subgroups Γ of $GL(2,\mathbb{F}_q)$. Here we look only at complex-valued modular forms. Of course, you could also think about finite field-valued modular forms. Perhaps it can be argued that the most natural analogue of the modular group $SL(2,\mathbb{Z})$ acting on the Poincaré upper half-plane is $GL(2,\mathbb{F}_p)$ acting on H_q, assuming $q = p^r$, with $r > 1$. One can also consider more general subgroups as do Shaheen [583, 584], and Hamahata [245]. Shaheen [583] shows that we have analogues of some of the modular functions in Sect. 3.4 and the present section as well as analogues of the special functions of Sect. 3.2, such as the power function, the K- and k-Bessel functions, and the gamma function. See also Shaheen and Terras [585].

First we need a multiplicative character χ of $\mathbb{F}_q^*$. Suppose that g is a generator of the multiplicative group $\mathbb{F}_q^*$. The **multiplicative character** $\chi = \chi_a$ has the following form for integers a, b:

$$\chi_a(g^b) = e^{\frac{2\pi iab}{q-1}}, \quad 0 \le a,b \le q-2. \tag{3.91}$$

Then we define the **power function** associated to χ by

$$p_\chi(z) = \chi(\operatorname{Im} z).$$

As with the power function on the Poincaré upper half-plane H, $p_\chi(z)$ is an eigenfunction of all the adjacency operators A_a for finite upper half-plane graphs in formula (3.53).

Let Ψ be an **additive character** of $\mathbb{F}_q$. Then for some $b \in \mathbb{F}_q$, we have $\Psi = \Psi_b$ where

$$\Psi_b(u) = e^{\frac{2\pi i Tr(bu)}{p}}, \quad \text{for } b,u \in \mathbb{F}_q. \tag{3.92}$$

Here if $q = p^r$, the **trace** in the exponent is

$$Tr(u) = Tr_{\mathbb{F}_q/\mathbb{F}_p}(u) = u + u^p + u^{p^2} + \cdots + u^{p^{r-1}}. \tag{3.93}$$

Let Ψ be an additive character and χ a multiplicative character of $\mathbb{F}_q$. Define the **Gauss sum** by

$$\Gamma(\chi,\Psi) = \Gamma_q(\chi,\Psi) = \sum_{t\in\mathbb{F}_q^*} \chi(t)\Psi(t). \tag{3.94}$$

The Gauss sum is analogous to the gamma function of formula (1.27).

See Ireland and Rosen [317] for more information on Gauss sums. In particular, they show that, if χ and Ψ_b are not trivial, then

$$\Gamma(\chi,\Psi_b)=\chi(b^{-1})\Gamma(\chi,\Psi_1). \tag{3.95}$$

Another basic result says that if χ and Ψ are not trivial, then

$$\left|\Gamma_q(\chi,\Psi)\right|=\sqrt{q}. \tag{3.96}$$

The **Kloosterman sum** for $a,b\in\mathbb{F}_q^*$ is defined by

$$K_\Psi(\chi|a,b)=\sum_{t\in\mathbb{F}_q^*}\chi(t)\Psi\left(at+bt^{-1}\right). \tag{3.97}$$

The Kloosterman sum is analogous to the K-Bessel function of formula (3.17). These sums have great importance in number theory. See Sarnak [557] who discusses the connection with modular forms and "Kloostermania."

For a multiplicative character χ of $\mathbb{F}_q^*$ and an additive character Ψ of $\mathbb{F}_q$, if $z\in H_q$, define the **k-Bessel function** by

$$k(z|\chi,\Psi)=\sum_{u\in\mathbb{F}_q}\chi\left(\operatorname{Im}\left(\frac{-1}{z+u}\right)\right)\Psi(u).$$

This is an analogue of the k-Bessel function in formula (3.16) and, by definition, $k(z)=k(z|\chi,\Psi)$ is an eigenfunction of all the adjacency operators A_a in formula (3.53) if $k(z)$ does not vanish identically on H_q.

Nancy Celniker et al. [82] find a finite analogue of formula (3.18) relating k- and K-Bessel functions. First we must recall the definition of the **quadratic residue symbol** ε_q by

$$\varepsilon_q(y)=\begin{cases}\;\;1, \text{ if } y \text{ is a square in } \mathbb{F}_q^*;\\ -1, \text{ if } y \text{ is not a square in } \mathbb{F}_q^*;\\ \;\;0,\ y=0.\end{cases} \tag{3.98}$$

We saw this symbol before in formula (1.33).

For a nontrivial multiplicative character χ of $\mathbb{F}_q^*$ and a nontrivial additive character Ψ_a of $\mathbb{F}_q$ and $z\in H_q$, we have the **relation between k-Bessel functions on H_q and Kloosterman sums:**

$$\Gamma_q(\chi,\Psi_a)k(z|\chi,\Psi_a)=g_a\chi(y)\Psi_a(-x)K_{\Psi_a}\left(\chi\varepsilon_q\middle|-\delta y^2,-\frac{1}{4}\right), \tag{3.99}$$

where $g_a=\sum_{u\in\mathbb{F}_q}\Psi_a(u^2)$ is a Gauss sum and the Kloosterman sum K is defined by formula (3.97).

Why do we call g_a a Gauss sum? It is an exercise to show that for $a\neq 0$

$$g_a = \sum_{u \in \mathbb{F}_q} \Psi_a(u^2) = \sum_{u \in \mathbb{F}_q} (1 + \varepsilon_q(u))\, \Psi_a(u) = \varepsilon_q\left(a^{-1}\right) \Gamma_q(\varepsilon_q, \Psi_1). \tag{3.100}$$

Now we define an Eisenstein series that is analogous to the Maass Eisenstein series defined in formula (3.81). Let Γ be a subgroup of $GL(2, \mathbb{F}_q)$ and let χ be a multiplicative character on $\mathbb{F}_q^*$. Define the finite **Maass Eisenstein "series"** for $z \in H_q$ as

$$E_{\chi,\Gamma}(z) = \sum_{\gamma \in \Gamma} \chi(\mathrm{Im}(\gamma z)). \tag{3.101}$$

Note that we do not sum over $\Gamma_\infty \backslash \Gamma$, where Γ_∞ is the subgroup of Γ consisting of matrices $\begin{pmatrix} a & b \\ 0 & a \end{pmatrix} \in \Gamma$. We do not need to do this, since Γ_∞ is finite in this case and thus just contributes a factor of $|\Gamma_\infty|$.

By definition, $E_{\chi,\Gamma}$ is an eigenfunction of the adjacency operators in formula (3.53) for finite upper half-plane graphs as long as it does not vanish identically on H_q. We found in Shaheen and Terras [585] that for $q = p^r$, with $q > 2$ and $r > 2$, $E_{\chi,GL(2,\mathbb{F}_p)}$ is not identically zero if and only if $\chi = \chi_a$ when a is a multiple of $(p-1)$. This last condition is equivalent to saying that $\chi_a|_{\mathbb{F}_p^*} = 1$. For $f, g : H_q \to \mathbb{C}$, define the inner product

$$(f, g) = \sum_{z \in H_q} f(z)\overline{g(z)}.$$

Computing $\left(E_{\chi_a}, E_{\chi_b}\right)$ allows one to show that the Eisenstein series are also not identically zero when $r = 2$ and $\chi = \chi_a$ when a is a multiple of $(p-1)$. It also allows one to find out when these Eisenstein series are orthogonal. We leave this as an exercise for the reader.

One the main results of Shaheen and Terras [585] is a finite analogue of Exercise 3.5.4 and says that the **Fourier expansion of the Eisenstein series** $E_{\chi,GL(2,\mathbb{F}_q)}(z)$ is given by

$$\begin{aligned} \frac{1}{p(p-1)^2} E_{\chi,GL(2,\mathbb{F}_p)}(z) &= \chi(y) + \frac{p}{q} \frac{\Gamma_q(\varepsilon_q, \Psi_1)\Gamma_q(\varepsilon_q \chi, \Psi_1)}{\Gamma_q(\chi, \Psi_1)} \chi^{-1}(-\delta y)\varepsilon_q(-\delta) \\ &+ \frac{p}{q} \frac{\Gamma_q(\varepsilon_q, \Psi_1)}{\Gamma_q(\chi, \Psi_1)} \chi(y) \sum_{\substack{b \in \mathbb{F}_q^* \\ Tr(b)=0}} \chi(b)\varepsilon_q(b) K_{\Psi_b}\left(\varepsilon_q \chi \,\middle|\, -\delta y^2, -\frac{1}{4}\right) \Psi_b(-x). \end{aligned} \tag{3.102}$$

Shaheen and Terras also prove the analogous result for $\Gamma = SL(2, \mathbb{F}_p)$. The key ingredient in the proof is Poisson summation for the subgroup $\mathbb{F}_p$ of $\mathbb{F}_q$.

Many questions remain on this topic. Are there finite analogues of cusp forms? Can we obtain them by averaging over Γ the spherical functions associated to discrete series representations of G? That is take the Soto–Andrade spherical function of formula (3.56) and average it over Γ. We leave such questions to the reader.

The paper [245] of Hamahata looks at finite analogues of Hilbert modular forms (to be discussed in Volume II, i.e., [667]) for a group $\Gamma \subset GL(2,\mathbb{F}_{p^2})$, where $\gamma \in GL(2,\mathbb{F}_{p^2})$ acts on $(z,z') \in H_p \times H_p$ by $\gamma(z,z') = (\gamma z, \gamma' z')$. Here γ' is the matrix obtained from γ by taking conjugates of all entries, where the conjugate on elements of $\mathbb{F}_{p^2}$ is defined as follows if $\mathbb{F}_{p^2} = \mathbb{F}_p\left(\sqrt{\delta}\right)$, where δ is a nonsquare in $\mathbb{F}_p$:

$$\left(x+y\sqrt{\delta}\right)' = x - y\sqrt{\delta}, \text{ for } x,y \in \mathbb{F}_p.$$

3.6 Modular Forms and Dirichlet Series. Hecke Theory and Generalizations

> ... Hecke took up the subject of modular functions and put it back into number theory where it had always belonged....
>
> —From Weil [729, Vol. III, p. 301].

3.6.1 *Dirichlet Series Corresponding to Holomorphic Modular Forms: The Hecke Correspondence*

We saw in the proof of Theorem 1.4.1 of Sect. 1.4 that the Mellin transform allows one to deduce the basic properties of the Epstein zeta function from those of the theta function. This is only a special case of a very general theory for which the boundaries have not yet been seen.

In the 1930s, Hecke looked at the following situation (see Hecke [258, pp. 591–626, 627–643, 644–671, 672–707, etc.] and [259]). Suppose that f is a classical holomorphic modular form (see Sect. 3.4); i.e., $f \in \mathcal{M}(SL(2,\mathbb{Z}),k)$ with Fourier expansion

$$f(z) = \sum_{n\geq 0} a_n \exp(2\pi i n z), \tag{3.103}$$

and consider the Mellin transform

$$\Lambda_f(s) = \int_0^\infty (f(iy) - a_0) y^{s-1} dy. \tag{3.104}$$

Then from Euler's formula for the gamma function, we know that for Re $x > 0$ and Re $s > 0$,

$$x^{-s}\Gamma(s) = \int_0^\infty y^{s-1}e^{-xy}dy$$

(see Lebedev [401, Chap. 1] for a discussion of $\Gamma(s)$). It follows that

$$\Lambda_f(s) = (2\pi)^{-s}\Gamma(s)L_f(s) \quad \text{with } L_f(s) = \sum_{n\geq 1} a_n n^{-s}. \tag{3.105}$$

We define $L_f(s)$ to be the **L-series corresponding to the modular form** $f(z)$. The L-series in Eq. (3.105) converges for Re $s > k$, by Theorem 3.4.4 of Sect. 3.4.

Exercise 3.6.1 (Analytic Continuation and Functional Equation of L-Series Corresponding to Holomorphic Modular Forms).

(a) Show that $f \in \mathcal{M}(SL(2,\mathbb{Z}),k)$ implies that $\Lambda_f(s)$ defined by Eq. (3.104) can be continued to a meromorphic function of $s \in \mathbb{C}$ so that

$$\Lambda_f(s) + a_0\left(\frac{1}{s} + \frac{i^k}{k-s}\right)$$

is **entire and bounded in vertical strips** (**EBV** for short).

(b) Show that, moreover, $\Lambda_f(s)$ satisfies the **functional equation**

$$\Lambda_f(s) = i^k\Lambda_f(k-s).$$

Hint. *Imitate the proof of Theorem 1.4.1 of Sect. 1.4. That is, split up the Mellin transform representing $\Lambda_f(s)$ into*

$$\int_0^1 + \int_1^\infty .$$

Then use the transformation formula $f(z) = z^{-k}f(-1/z)$ to rewrite the integral over $[0,1]$. The result is that, as before, the L-series can be evaluated by an ***incomplete gamma expansion****:*

$$\Lambda_f(s) = a_0\left(\frac{i^k}{s-k} - \frac{1}{s}\right) + \sum_{n\geq 1} a_n\left(G(s,2\pi n) + G(k-s,2\pi n)i^k\right), \tag{3.106}$$

with the incomplete gamma function

$$G(s,x) = \int_1^\infty t^{s-1}\exp(-xt)dt \quad \text{for Re } x > 0,$$

as in Theorem 1.4.1 of Sect. 1.4. The exercise follows from this expansion, using the exponential decay of $G(s,x)$, as x approaches infinity (see Exercise 1.4.7 of Sect. 1.4).

Hecke noticed that Mellin inversion (see Sect. 1.4) allows the argument of Exercise 3.6.1 to be turned around.

Exercise 3.6.2 (A Converse Result in Hecke Theory). Suppose that $f(z)$, with Fourier expansion in formula (3.103), is given and consider the function $\Lambda_f(s)$ defined by Eqs. (3.104) and (3.105). Suppose that $\Lambda_f(s)$ has the functional equation

$$\Lambda_f(s) = i^k \Lambda_f(k-s).$$

and that

$$\Lambda_f(s) + a_0\left(\frac{1}{s} + \frac{i^k}{k-s}\right)$$

is entire and bounded in vertical strips (EBV). Prove that then $f(z)$ is a holomorphic modular form of weight k for $SL(2,\mathbb{Z})$; i.e., that f lies in $\mathcal{M}(SL(2,\mathbb{Z}),k)$.

Hint. *Note that $z \mapsto z+1$ and $z \mapsto -1/z$ generate $SL(2,\mathbb{Z})/\{\pm I\} = PSL(2,\mathbb{Z})$. Thus, it suffices for us to show that $f(z) = z^{-k} f(-1/z)$. Use Mellin inversion to write*

$$f(iy) - a_0 = \frac{1}{2\pi i}\int_{\mathrm{Re}\, s = c} y^{-s}\Lambda_f(s)ds,$$

where c is sufficiently large for the absolute convergence of Λ_f. Push the line of integration to the left and pick up residues at $s = k$ and $s = 0$, to obtain

$$f(iy) - a_0 = \frac{1}{2\pi i}\int_{\mathrm{Re}\, s = k-c} y^{-s}\Lambda_f(s)ds + i^k y^{-k} a_0 - a_0.$$

Then use $\Lambda_f(k-s) = \Lambda_f(s) i^k$, to see that $f(iy) = (i/y)^k f(i/y)$ for all $y > 0$. To complete the exercise, note that if two holomorphic functions on H agree on a set with an accumulation point, they must agree on all of H.

Some references for these exercises are Hecke [258,259], Lang [391], Ogg [505], and Shimura [589]. Note that more general groups than $SL(2,\mathbb{Z})$ can also be treated in this way. The following theorem follows from the exercises.

Theorem 3.6.1 (Hecke's Correspondence for Holomorphic Modular Forms for SL(2, $\mathbb{Z}$)). *Suppose that $f(z)$ has the Fourier expansion given in formula (3.103). Suppose that the corresponding Dirichlet series is given by formulas (3.104) and (3.105). Then we have the equivalence*

$$\begin{aligned} f &\in \mathcal{M}(SL(2,\mathbb{Z}),k) \\ &\Longleftrightarrow \Lambda_f(s) + a_0\left(\frac{1}{s} + \frac{i^k}{k-s}\right) \text{ is EBV with } \Lambda_f(s) = i^k\Lambda_f(k-s). \end{aligned}$$

Here EBV means entire and bounded in vertical strips.

Example 3.6.1. The Dirichlet series corresponding to the classical Eisenstein series G_k from Sect. 3.4 is

$$\sum_{n\geq 1} \sigma_{k-1}(n)n^{-s} = \zeta(s)\zeta(s+1-k). \tag{3.107}$$

The functional equation can actually be deduced from that of the Riemann zeta function, using the duplication formula for the gamma function and the functional equation of the gamma function.

Exercise 3.6.3. Prove the assertions made in the preceding example.

3.6.2 Dirichlet Series Corresponding to Maass Waveforms

From our point of view, the next logical question was answered in the 1940s by Maass [439]. This question is: What is the analogue of Theorem 3.6.1 for Maass waveforms? Another reference for the answer is Maass [437].

In Hecke's case we needed to know the Mellin transform

$$\int_0^\infty e^{-y}y^{s-1}dy = \Gamma(s) \quad \text{for} \quad \mathrm{Re}\, s > 0. \tag{3.108}$$

In Maass's case we need to know the Mellin transform

$$\int_0^\infty K_r(y)y^{s-1}dy = 2^{s-2}\Gamma\left(\frac{s+r}{2}\right)\Gamma\left(\frac{s-r}{2}\right), \quad \text{if} \quad \mathrm{Re}\, s > |\mathrm{Re}\, r|. \tag{3.109}$$

Exercise 3.6.4. Prove formula (3.109). This was part (b) of Exercise 3.2.7 in Sect. 3.2.

Hint. *Use the integral formula in Exercise 3.2.1 of Sect. 3.2 for the K-Bessel function to see that*

$$\int_{y>0} K_r(y)y^{s-1}dy = \frac{1}{2}\int_{y>0}\int_{t>0} y^{s-1}t^{r-1}\exp\left[-\frac{1}{2}y\left(t+t^{-1}\right)\right] dt\, dy.$$

Then make the substitution $u = yt$, $v = y/t$, to complete the proof.

Let f be a Maass waveform; i.e., $f \in \mathcal{N}(SL(2,\mathbb{Z}), r(r-1))$, with Fourier expansion as given in Exercise 3.5.3 of Sect. 3.5:

$$f(z) = ay^r + by^{1-r} + \sum_{n\neq 0} a_n y^{\frac{1}{2}} K_{r-\frac{1}{2}}(2\pi|n|y) exp(2\pi inx). \tag{3.110}$$

Consider the Mellin transform:

$$W_f(s) = \int_0^\infty \left(f(iy) - ay^r - by^{1-r}\right) y^{s-1}\, dy. \tag{3.111}$$

Then, Exercise 3.6.4 shows that

$$W_f(s) = 2^{-5/2}\pi^{-(s+\frac{1}{2})}\Gamma\left(\frac{s-r+1}{2}\right)\Gamma\left(\frac{s+r}{2}\right)L_f(s), \tag{3.112}$$

with

$$L_f(s) = \sum_{n\neq 0} a_n |n|^{-(s+\frac{1}{2})}. \tag{3.113}$$

We define $L_f(s)$ to be the ***L*-series corresponding to the Maass waveform** $f \in \mathcal{N}(SL(2,\mathbb{Z}), r(r-1))$. This series converges by Theorem 3.5.2 of Sect. 3.5 when $\mathrm{Re}\, s > |\mathrm{Re}\, r|$, if f is not a cusp form, and for $\mathrm{Re}\, s > 1$, if f is a cusp form.

Exercise 3.6.5 (Analytic Continuation and Functional Equation of L-Series Corresponding to Maass Waveforms). Define the *higher-dimensional incomplete gamma functions* $G(s_1, s_2; a)$ by

$$G(s_1,s_2;a) = \iint_{uv\geq 1} u^{s_1-1}v^{s_2-1}e^{-a(u+v)}\, du\, dv, \quad \text{for } \mathrm{Re}\, s_i > 0 \text{ and } \mathrm{Re}\, a > 0.$$

(a) Use the same trick as in Exercise 3.6.1 to show that $W_f(s)$ in Eq. (3.112) has the following incomplete gamma expansion, assuming $f \in \mathcal{N}(SL(2,\mathbb{Z}), r(r-1))$:

$$W_f(s) = a\left(\frac{1}{s-r} + \frac{1}{-s-r}\right) + b\left(\frac{1}{s+r-1} + \frac{1}{-s+r-1}\right)$$
$$+ \sum_{n\neq 0} a_n \left\{ G\left(\frac{s+r}{2}, \frac{s-r+1}{2}; \pi|n|\right) + G\left(\frac{-s+r}{2}, \frac{1-s-r}{2}; \pi|n|\right)\right\}.$$

(b) Show that the incomplete gamma function $G(s_1, s_2; a)$ dies off exponentially as a approaches infinity.

(c) Deduce that $W_f(s)$ continues to a meromorphic function of s such that

$$W_f(s) - a\left(\frac{1}{s-r} + \frac{1}{-s-r}\right) - b\left(\frac{1}{s+r-1} + \frac{1}{-s+r-1}\right)$$

is EBV, with functional equation $W_f(s) = W_f(-s)$.

Hint. *To prove (a) use the same substitution that appeared in Exercise 3.6.4.*

Now we want to argue that Mellin inversion implies a converse to Exercise 3.6.5, analogous to Hecke's converse result in Theorem 3.6.1. However, in Exercise 3.6.2, we use a fact about holomorphic functions to show that $f(iy) = (i/y)^k f(i/y)$ for all $y > 0$ implies that $f(z) = z^{-k} f(-1/z)$ for all z in H. But real analytic functions $f(z)$ and $g(z)$ which coincide on $\operatorname{Re} z = 0$ are not necessarily equal. For an example, let $f(z) = \exp(iy)$ and $g(z) = \exp(x + iy)$. Thus we need the following exercise from Maass [439].

Exercise 3.6.6 (Facts About Eigenfunctions of the Laplacian). Suppose that f is an eigenfunction of the non-Euclidean Laplacian; i.e., $\Delta f = \lambda f$, and suppose, in addition, that $f|_{x=0} = f_x|_{x=0} = 0$ for all $y > 0$. Here $f_x = \partial f/\partial x$. Show that f must be identically zero on H.

Hint. *We know that $f(z) = \Sigma_{n\geq 0} c_n(y) x^n$. Thus $c_n(y)$ satisfies the recursion obtained from*

$$0 = \Delta f - \lambda f = \sum_{n\geq 0} \left(n(n-1) y^2 c_n(y) x^{n-2} + y^2 c_n''(y) x^n - \lambda c_n(y) x^n \right).$$

The initial conditions then imply that the $c_n(y)$ are identically zero. Note that without the hypothesis that $\partial f/\partial x|_{x=0} = 0$, we could only prove that $c_{2n}(y) = 0$.

In order to obtain the second initial condition for the function $f(z) - f(-1/z)$, we need the **second L-series corresponding to the Maass waveform** $f \in \mathcal{N}(SL(2,\mathbb{Z}), r(r-1))$:

$$L_{f_x}(s) = \sum_{n\neq 0} (2\pi i n) a_n |n|^{-(s+\frac{1}{2})}. \tag{3.114}$$

This second Dirichlet series is also seen to be necessary because the first series L_f, from Eq. (3.112), will vanish identically if f is an odd function of x; i.e., if $f(x + iy) = -f(-x + iy)$. For then $a_n = -a_{-n}$. And Table 3.5 of Sect. 3.5 as well as the tables of Booker et al. [48] give some evidence for the conjecture that the Dirichlet and Neumann problems have different eigenvalues. A cusp form f has the decomposition $f = u + v$, where u is a solution of the Dirichlet problem and v is a solution of the Neumann problem. So, probably, either $f = u$ or $f = v$; i.e., f is either even or odd, as a function of x. If $f(z)$ denotes a Maass waveform solving the Dirichlet problem, we see that the L-series $L_f(s) = 0$, and we must consider $L_{f_x}(s)$ to obtain a nonzero series.

Using the differentiated modular form is an old trick in number theory. One needs it to obtain the analytic continuation of Dirichlet L-series:

$$\sum_{n\geq 1} \chi(n) n^{-s}$$

when $\chi(-1) = -\chi(1)$ and χ is a multiplicative character on the group of units in $\mathbb{Z}/m\mathbb{Z}$, viewed as a function on $\mathbb{Z}$ by writing $\chi(a) = 0$ if $\gcd(a,m) \neq 1$. A reference for the analytic continuation of these L-series is Davenport [120]. Hecke generalized this to grossencharacters for algebraic number fields (see Hecke [258, pp. 215–234, 249–289]).

Exercise 3.6.7 (Analytic Continuation and Functional Equation for the Second L-Series Corresponding to a Maass Waveform). Note that $f(-1/z) = f(z)$ implies that $f_x(-1/z) = z^2 f_x(z)$; so f_x has weight 2. Imitate the argument in Exercise 3.6.5 to obtain the analytic continuation and functional equation of L_{f_x}.

The following theorem can now be proved as an Exercise.

Theorem 3.6.2 (The Maass Correspondence Between Maass Waveforms and Dirichlet Series). *Suppose that $f(z)$ has the Fourier expansion given by formula (3.110). Then $f \in \mathcal{N}(SL(2,\mathbb{Z}), r(r-1))$ is equivalent to the following assertions about the Dirichlet series defined in formulas (3.112)–(3.114):*

(1)

$$W_f(s) - a\left(\frac{1}{s-r} + \frac{1}{-s-r}\right) - b\left(\frac{1}{s+r-1} + \frac{1}{-s+r-1}\right)$$

is EBV with functional equation $W_f(s) = W_f(-s)$, and

(2) W_{f_x} has the analogous properties (see Exercise 3.6.7).

Exercise 3.6.8. Complete the proof of Theorem 3.6.2.

Maass actually proved a result like Theorem 3.6.2 for functions $f(z)$ satisfying more general differential equations and invariant under more general discrete subgroups of $SL(2,\mathbb{R})$.

Expansions of Dirichlet series in series of higher-dimensional incomplete gamma functions have been studied by many people. They arise whenever there are several gamma factors in the functional equation of the Dirichlet series (see Goldfeld and Viola [222], Lavrik [399], and Terras [662, 663]).

Example 3.6.2. The Dirichlet series corresponding to the nonholomorphic Eisenstein series

$$E_r^*(z) = \pi^{-r}\Gamma(r)\zeta(2r)E_r(z)$$

is

$$4\sum_{n\geq 1}\sigma_{1-2r}(n)n^{r-\frac{1}{2}}n^{-s-\frac{1}{2}} = 4\zeta(s-r+1)\zeta(s+r). \tag{3.115}$$

Here again the functional equations can be deduced from that of the Riemann zeta function.

Exercise 3.6.9. Prove the assertions made in the preceding example.

3.6.3 Remarks on Extensions to Congruence Subgroups

One can use slight extensions of Hecke's theorem to derive the product expansion of Δ and θ in Sect. 3.4 (see Ogg [505, Chap. I, pp. 43–46], [506]). However, many important applications require a major extension of the theory, in particular, to congruence subgroups such as $\Gamma(N)$, which is defined after formula (3.51). Hecke and Maass both considered extensions of the theory to congruence subgroups. A simple example to explain why Hecke was interested in congruence subgroups can be found in Hecke's lectures, which were given at the Institute for Advanced Study at Princeton in 1938 (see Hecke [259, p. 43]). There it is noted that the zeta function for the imaginary quadratic field $\mathbb{Q}(\sqrt{-7})$ has as its inverse Mellin transform a theta function which is a modular form for $\Gamma_0(7) =$ the congruence subgroup of $SL(2,\mathbb{Z})$ consisting of matrices whose lower left entry is divisible by 7. Hecke also made connections with the representations of the finite group $SL(2,\mathbb{Z}/n\mathbb{Z})$ (see also Eichler [151]).

In 1967, Weil extended Hecke's correspondence between modular forms and Dirichlet series to congruence subgroups in a different way (see Weil [729, Vol. 3, pp. 165–172]). It might be interesting to connect the two points of view. In fact, Weil extended Hecke's theory to the subgroup $\Gamma_0(N)$ consisting of all elements of $SL(2,\mathbb{Z})$ for which the lower left entry is divisible by N. Here Weil discovered that the converse theorem requires more than one functional equation. Because the groups involved do not even contain the inversion $z \mapsto -1/z$, one has to add the matrix

$$\begin{pmatrix} 0 & -1 \\ N & 0 \end{pmatrix},$$

forming a larger group than $\Gamma_0(N)$. The resulting group will have more than two generators, in general. Weil finds that to make $f(z)$, with Fourier coefficients a_n, a modular form for $\Gamma_0(N)$, one requires the functional equations of the L-functions

$$\sum_{n\geq 1} \chi(n) a_n n^{-s}$$

for sufficiently many primitive characters χ of the unit group $(\mathbb{Z}/m\mathbb{Z})^* = \{b \bmod m \mid \gcd(b,m) = 1\}$. Here one considers only m that are relatively prime to N. The functional equations of these L-functions involve Gauss sums. Other references for Weil's result are Diamond and Shurman [133], Koblitz [359], Lang [391], Ogg [505], Razar [539], and Shimura [589].

Weil's result involves explicit formulas for elements of the congruence subgroup. From the point of view of harmonic analysis, one might expect that harmonic analysis on $\Gamma(N)\backslash H$ should be a combination of analysis on $\Gamma(1)\backslash H$ and that on $\Gamma(1)/\Gamma(N)$. The latter quotient is isomorphic to $SL(2,\mathbb{Z}/N\mathbb{Z})$. The principal series representations of the finite special linear group are related to characters of the group $(\mathbb{Z}/N\mathbb{Z})^*$ (see Kirillov [353, p. 267]). This sort of thinking would lead one to expect a result similar to Weil's but involving L-functions formed with characters $\bmod N$ itself, rather than $\bmod m$, with $\gcd(m,N) = 1$ (see Razar [539]).

Maass [439] actually obtains the correspondence between nonholomorphic modular forms for congruence subgroups and Dirichlet series with functional equations involving two gamma factors. This allows Maass to prove the existence of cuspidal Maass waveforms for congruence subgroups, distinct from $SL(2,\mathbb{Z})$ itself, by using the fact that the functional equations of Hecke L-functions with grossencharacters for real quadratic fields involve two gamma factors. These are the CM Maass forms of Hejhal and Strombergsson [275].

Jacquet and Langlands [323] give an adelic view of Hecke theory. See also the collection of papers introducing the Langlands program edited by Bernstein and Gelbart [37], as well as Gelbart [200] and Weil [728], plus Borel and Casselman [52]. Another reference is Winnie Li [416–419], who has obtained an adelic version of Hecke theory closer to that envisioned above from the point of view of harmonic analysis.

Hecke theory has implications for certain conjectures of number theory and geometry. Number theorists have been interested in Weil's result for weight 1 modular forms because of the **Artin conjecture** that Artin L-functions associated to representations of Galois groups of number fields are entire provided the representation is nontrivial and irreducible. References for Artin L-functions include Heilbronn's article in Cassels and Fröhlich [79, pp. 204–230], Lang [388], and Stark [632]. This conjecture has implications for the Dedekind zeta functions of number fields. For example, it implies that whenever a number field F is contained in a larger number field K, the quotient of Dedekind zeta functions given by $\zeta_K(s)/\zeta_F(s)$ is an entire function of s. This result is known if K/F is Galois or normal. Putting results of Weil, Langlands, Deligne, and Serre together, one sees that (modulo the Artin conjecture) certain forms of weight 1 for $\Gamma_0(N)$ correspond to Artin L-functions for irreducible two-dimensional representations of the Galois group of the algebraic closure of the rationals. Because Artin L-functions have functional equations involving many gamma factors, there is a general feeling that these L-functions must somehow correspond to automorphic forms for $GL(n)$. This sort of feeling is part of the "Langlands philosophy." References for some of these things are Borel and Casselman [52], Serre [577], Tate [658], and Weil [729, Vol. III, notes especially]. Langlands [394] and Tunnell [686] have proved some new cases of the Artin conjecture for two-dimensional representations of Galois groups using base change and twisted versions of the trace formula.

There is also a geometric conjecture related to Hecke theory. This is the **conjecture of Taniyama, Shimura, and Weil** which says that any elliptic curve over the rationals has a zeta function coming from a modular form of weight 2. It can be shown that modular forms of weight 2 give rise to L-series which are zeta functions of elliptic curves or abelian varieties. The conjecture of Taniyama, Shimura, and Weil has been proved and is now called the **modularity theorem**. As part of his proof of Fermat's last theorem, Andrew Wiles proved the modularity theorem for a large set of elliptic curves, with a key ingredient from his joint work with Richard Taylor. See [661, 740]. The complete modularity theorem was proved by Breuil et al. [66]. See Diamond and Shurman for an introduction to this subject with countless versions of the modularity theorem stated. See Weil [729, Vol. III, notes] for some discussion of the Taniyama, Shimura,Weil conjecture.

See Lang's article [392] about the naming of the conjecture. There are also many open questions on the Hasse–Weil zeta functions of algebraic varieties (even elliptic curves). References include Katz [344, pp. 300–301] and Swinnerton-Dyer [650].

Exercise 3.6.10 (Operators That Raise the Level).

(a) Suppose that χ is a multiplicative character $(\operatorname{mod} m)$; i.e.,

$$\chi : (\mathbb{Z}/m\mathbb{Z})^* \to \mathbb{T}, \quad \text{a homormorphism of multiplicative groups,}$$

where $(\mathbb{Z}/m\mathbb{Z})^*$ is the multiplicative group of integers $a \operatorname{mod} m$, with a relatively prime to m, and $\mathbb{T}$ is the multiplicative group of complex numbers of norm one. We extend χ to a function on $\mathbb{Z}^+$ by setting $\chi(n) = 0$ if n and m are not relatively prime. Suppose that $f \in \mathcal{M}(SL(2,\mathbb{Z}), k)$. Define

$$L_\chi f(z) = \frac{1}{m} \sum_{u,v (\operatorname{mod} m)} \chi(u) \exp(-2\pi i uv/m) f(z + v/m).$$

Show that $L_\chi f$ is a modular form for the congruence group $\Gamma_0(m^2)$.

(b) Suppose that a is relatively prime to m and a' is an integer such that $aa' \equiv 1 (\operatorname{mod} m)$. Let R_a be a matrix in $SL(2,\mathbb{Z})$ such that

$$R_a \equiv \begin{pmatrix} a' & 0 \\ 0 & a \end{pmatrix} \ (\operatorname{mod} N).$$

R_a is not well defined, but the coset of R_a in $\Gamma(N)\backslash SL(2,\mathbb{Z})$ is well defined. Set $N = m^2$ and show that

$$L_\chi f|_{R_a} = \chi^2(a) L_\chi f \quad \text{for } f \text{ as in part (a).}$$

Here the **slash operator** is defined by

$$g|_\gamma(z) = (cz+d)^{-k} g\left(\frac{az+b}{cz+d}\right), \quad \text{if}$$

$$\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL(2,\mathbb{Z}), \ g \text{ of weight } k.$$

(c) Suppose that $f(z)$ is as in part (a) and $f(z) = \sum_{n\geq 0} a_n \exp(2\pi i n z)$. Show that

$$L_\chi f(z) = \sum_{n\geq 0} a_n \chi(n) \exp(2\pi i n z).$$

This leads to the twisted L-function studied by Weil.

(d) Suppose that $f(z)$ is as in part (a). Let χ be a multiplicative character $\operatorname{mod} m$. Define

$$f_\chi(z) = \sum_{\substack{a (\mathrm{mod}\, m)\\ \gcd(a,m)=1}} \chi(a) f(z + a/m).$$

Show that f_χ is a modular form of weight k for $\Gamma_0(m^2)$ and that, using the notation of part (b), $f_\chi|_{R_a} = \chi^{-2}(a) f_\chi$.

(e) Can you connect the operators in parts (a) and (d)?

Hint. *Parts (a)–(c) can be found in Ogg [505, Chap. IV, pp. 37–38]. Part (d) comes from Eichler [151, pp. 147–149]. See also Razar [539].*

It will help to note that there is a chain of subgroups of $\Gamma = SL(2,\mathbb{Z})$ *given by*

$$\Gamma(N) \subset \Gamma_1(N) \subset \Gamma_0(n) \subset \Gamma,$$

where $\Gamma_1(N) = \left\{ \gamma \in \Gamma \,\middle|\, \gamma \equiv \begin{pmatrix} 1 & * \\ 0 & 1 \end{pmatrix} \mathrm{mod}\, N \right\}$. *The quotients* $\Gamma_1(N)/\Gamma(N)$ *and* $\Gamma_0(N)/\Gamma_1(N)$*are abelian; the first isomorphic to the additive group* $\mathbb{Z}/N\mathbb{Z}$ *and the second isomorphic to the multiplicative group* $(\mathbb{Z}/N\mathbb{Z})^*$.

Note also that [as in part (a) of Exercise 3.6.14 which follows], if γ *is a matrix of determinant N, then transforming* $f(z)$ *by* γ *using the slash operator*

$$f|_\gamma(z) = N^{k/2}(cz+d)^{-k} f((az+b)/(cz+d)) \quad \text{for } \gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

sends a form for $SL(2,\mathbb{Z})$ *of weight k to a form for* $\Gamma(N)$ *of weight k.*

Exercises 3.6.10 and 3.6.14 which follows, give a large number of examples of modular forms for congruence subgroups. Classically, examples of the form $f(mz)$ arose in the theory of equations for $f(mz)$ in terms of $f(z)$, for example, when $f(z)$ is a modular function, such as $j(z)$. See Shimura [589, pp. 109–110] for a discussion of the modular polynomial, also Weber [724, Vol. III].

3.6.4 Hecke Operators on Holomorphic Forms

Hecke also developed a theory that will determine what modular forms correspond to Dirichlet series with Euler products. The first Euler product to arise in number theory was that for the Riemann zeta function:

$$\zeta(s) = \prod_{p \text{ prime}} (1 - p^{-s})^{-1}, \ \mathrm{Re}\, s > 1 \tag{3.116}$$

(see Exercise 1.4.4 of Sect. 1.4). It is this result that governs the application of the Riemann zeta function to the study of the distribution of primes.

Around 1935, Hecke used Hecke operators to answer the question: Which modular forms correspond to L-series with Euler products? (see Hecke [258, pp. 577–707] and [259]). Such operators had already been considered by Hurwitz [311, pp. 163–188] and Mordell [476]. It turns out that Hecke operators are related to the analogue for p-adic groups of the spherical functions considered in Sect. 3.2 (see Gelbart [200, pp. 47 ff]). There is now a general theory of Hecke operators for a group G and a subgroup H. See Krieg [373] and Shimura [589].

As usual, we shall restrict ourselves to $SL(2,\mathbb{Z})$, although Hecke did not. For each positive integer n, let M_n denote the set of all 2×2 matrices of determinant n. Set $\Gamma = SL(2,\mathbb{Z})$. It is easy to see that M_n is a disjoint union:

$$M_n = \bigcup_{\substack{ad=n, d>0,\\ b(\operatorname{mod} d)}} \Gamma \begin{pmatrix} a & b \\ 0 & d \end{pmatrix}. \tag{3.117}$$

Exercise 3.6.11. Prove formula (3.117).

Hint. *It is clear that given a matrix $L \in M_n$, there exists a matrix $\gamma \in SL(2,\mathbb{Z})$ such that γL has zero for its lower left entry. Here we use the fact that a row vector of integers is the second row of a matrix in $SL(2,\mathbb{Z})$ if and only if the entries of the vector are relatively prime.*

Definition 3.6.1. The **Hecke operator** T_n is defined on $f \in \mathcal{M}(SL(2,\mathbb{Z}),k)$ by

$$T_n f(z) = n^{k-1} \sum_{\substack{ad=n, d>0,\\ b(\operatorname{mod} d)}} d^{-k} f\left(\frac{az+b}{d}\right). \tag{3.118}$$

Clearly the Hecke operator T_n also depends on k, the weight of the modular form $f(z)$ on which T_n acts. If $f(z)$ is a modular form of weight k, then the Hecke operator is actually independent of the choice of representatives for $SL(2,\mathbb{Z})\backslash M_n$. Thus we can consider the sum in Eq. (3.118) to be a sum over $SL(2,\mathbb{Z})\backslash M_n$. We will see that analogous sums arise in the theory of automorphic forms for $SL(n,\mathbb{Z})$ in Volume II [667]. Other references for Hecke operators are Apostol [12, pp. 120–138], Kilford [350], Koblitz [359], Krieg [373], Lang [391, Chaps. 2 and 3], Ogg [505], Serre [576], Shimura [589], and Stein [638].

Theorem 3.6.3 (Properties of Hecke Operators on Spaces of Holomorphic Modular Forms).

(1) The Hecke operator T_n defined by formula (3.118) has the property that if $f \in \mathcal{M}(SL(2,\mathbb{Z}),k)$, then $T_n f \in \mathcal{M}(SL(2,\mathbb{Z}),k)$. Moreover, if f is a cusp form, so is $T_n f$.

(2) Suppose that $f \in \mathcal{M}(SL(2,\mathbb{Z}),k)$ has Fourier expansion

$$f(z) = \sum_{m\geq 0} a_m \exp(2\pi i m z).$$

Then $T_n f$ has the Fourier expansion

$$T_n f(z) = \sum_{m\geq 0} b_m \exp(2\pi imz) \quad \textit{with } b_m = \sum_{d \mid \gcd(m,n)} d^{k-1} a_{mn/d^2}.$$

Here the sum defining b_m is over all positive divisors of the greatest common divisor of m and n.

(3) We have the ***multiplication formula***

$$T_n T_m = \sum_{d \mid \gcd(m,n)} d^{k-1} T_{mn/d^2}.$$

And the algebra generated by the Hecke operators $\{T_n,\ n \geq 1\}$ is commutative.

(4) Define the ***Petersson inner product*** *of two cusp forms $f, g \in \mathcal{S}(SL(2,\mathbb{Z}),k)$ by*

$$[f,g] = \int_{SL(2,\mathbb{Z})\backslash H} f(z)\overline{g(z)} y^{k-2}\, dx\, dy.$$

Then $[T_n f, g] = [f, T_n g]$. It follows that the Hecke operators can be simultaneously diagonalized on $\mathcal{M}(SL(2,\mathbb{Z}),k)$.

(5) Suppose that $f \in \mathcal{M}(SL(2,\mathbb{Z}),k)$ with

$$f(z) = \sum_{m\geq 0} a_m \exp(2\pi imz),$$

and $a_1 = 1$, satisfies $T_n f = c_n f$, for $n = 1,2,3,\ldots$. Then $a_n = c_n$, for all $n = 1,2,3,\ldots$. Moreover, the Dirichlet series associated to f has an ***Euler product*** *of the form*

$$L_f(s) = \sum_{n\geq 1} a_n n^{-s} = \prod_{p \text{ prime}} (1 - a_p p^{-s} + p^{k-1-2s})^{-1}.$$

Conversely, if $L_f(s) = \sum_{n\geq 1} a_n n^{-s}$ has such an Euler product and is known to correspond to a modular form f in $\mathcal{M}(SL(2,\mathbb{Z}),k)$, then f is an eigenfunction for all of the Hecke operators T_n.

(6) The ***eigenvalues*** *of the Hecke operators T_n are totally real algebraic numbers.*

Exercise 3.6.12. Prove the preceding theorem using the same sorts of arguments that we will give in the proof of Theorem 3.6.4, which concerns the analogue for Maass waveforms. Details can also be found in the references preceding Theorem 3.6.3. The only part of Theorem 3.6.3 that does not have an analogue in Theorem 3.6.4 is part (6). To see this, use the fact that the space of modular forms of weight k has a basis consisting of elements for which all the Fourier coefficients are rational (see Ogg [505, III–12]).

Example 3.6.3. The Eisenstein series G_k gives an example of an eigenfunction of the Hecke operators, since the corresponding Dirichlet series has a Euler product:

$$L_{G_k}(s) = \zeta(s)\zeta(s+1-k).$$

Example 3.6.4. The cusp form Δ of weight 12 must also be an eigenfunction of the Hecke operators, since it is unique, up to a constant multiple. Thus the corresponding Dirichlet series

$$L_{\Delta}(s) = \sum_{n \geq 1} \tau(n) n^{-s}$$

has an Euler product, a result first proved by Mordell [476] in 1917. Ramanujan had stated this result some years before Mordell's proof, as well as the Ramanujan conjecture (see Sect. 3.4).

Hardy [249, Chap. X] discussed the Ramanujan τ in his book on Ramanujan, beginning with the remark: "We may seem to be straying into one of the backwaters of mathematics, but the genesis of $\tau(n)$ as a coefficient in so fundamental a function compels us to treat it with respect." Weil considered this remark in his 1972 Columbia lectures (see Weil [729, Vol. 3, pp. 280–281]) as "a typical example of the deep gulf that separates number-theorists from an analyst like Hardy." This sort of mathematical name-calling is fun, but history clearly has a way of refuting all such assertions, including many made by the present author in this and earlier versions of this book.

More examples can be found in Hecke [258, 259]. In particular, paper no. 37 in Hecke's collected works [258] gives a list of Dirichlet series L_f, showing that the presence of an Euler product is neither necessary nor sufficient for the existence of infinitely many zeros on the line of symmetry $\operatorname{Re} s = k/2$.

Goldstein [225] has considered analogues of the Riemann hypothesis and Mertens conjecture for the Dirichlet series attached to cusp forms which are eigenfunctions for all the Hecke operators. Anderson [6] disproved the Mertens conjecture for these Dirichlet series corresponding to such cusp forms when the weight is sufficiently large. In 1985, Odlyzko and te Riele disproved the original Mertens conjecture (see [504]).

Bounds for the degree of the number fields containing the eigenvalues of Hecke operators are obtained in Rankin and Rushforth [535]. Naomi Jochnowitz [329] studied the algebra $T^{(k)}$ of Hecke operators acting on holomorphic cusp forms of weight k for $SL(2,\mathbb{Z})$. Then $T^{(k)} \otimes \mathbb{Q}$ is a direct product of totally real number fields whose dimension over $\mathbb{Q}$ equals the dimension of the space $\mathcal{S}(SL(2,\mathbb{Z}),k)$ over $\mathbb{C}$. Moreover, in all known cases the direct product is actually a single number field. For example, Hecke [258, p. 671] shows that $T^{(24)} \otimes \mathbb{Q} \simeq \mathbb{Q}(\sqrt{144169})$ and that $T^{(24)}$ is the unique suborder of index 24 in the ring of integers of this field. Jochnowitz shows that if N is any integer and k is sufficiently large then N divides the index of $T^{(k)}$ in the maximal order of $T^{(k)} \otimes \mathbb{Q}$.

The three collections of *Math. Reviews* in number theory (see [243, 414, 456]) contain references to many papers on Hecke operators, etc. For example, there are reviews of some of Hecke's papers (see Leveque [414, Vol. 2, pp. 448 ff]) as well as a formula of Petersson for $\zeta(3)$, involving his inner product and theta functions. Of course, one can also visit the *Math. Reviews* website or just search the Internet for Hecke operators. Wohlfahrt [743] generalizes the theory of Hecke operators to automorphic forms of real (rather than integral) weight. Rademacher [526] shows that T_{p^r} is essentially a Tchebychef polynomial of the second kind in T_p. Atkin and Lehner [16] complete the theory of Hecke operators for the congruence group $\Gamma_0(N)$. The problem was to find a satisfactory theory of operators for primes dividing the level of the congruence subgroup. Ihara [314] uses the Eichler–Selberg trace formula to connect Ramanujan's conjecture with the Weil conjectures on algebraic varieties—a project ultimately completed by Deligne, as we noted in Sect. 3.4.

3.6.5 Hecke Operators for Maass Wave Forms

The Hecke operators in the nonholomorphic case were developed by Maass [437, 439].

Definition 3.6.2. For a Maass waveform $f \in \mathcal{N}(SL(2,\mathbb{Z}), r(r-1))$, and $n = 1,2,3,\ldots$, define the **Hecke operator** T_n by

$$T_n f(z) = n^{-\frac{1}{2}} \sum_{\substack{ad=n, d>0 \\ b(\mathrm{mod}\, d)}} f\left(\frac{az+b}{d}\right). \tag{3.119}$$

The normalizing factor $n^{-\frac{1}{2}}$ is not consistent with formula (3.118), since our Maass waveforms have weight 0, but it has become standard in the Maass waveform literature.

Theorem 3.6.4 (Properties of Hecke Operators on Maass Waveforms).

(1) The Hecke operator T_n, defined by formula (3.119), has the property that if $f \in \mathcal{N}(SL(2,\mathbb{Z}), r(r-1))$, then $T_n f \in \mathcal{N}(SL(2,\mathbb{Z}), r(r-1))$. Moreover, if f is a cusp form, then so is $T_n f$. And the map $f \mapsto T_n f$ is linear.

(2) Suppose that $f \in \mathcal{N}(SL(2,\mathbb{Z}), r(r-1))$ has the Fourier expansion

$$f(z) = cy^r + c'y^{1-r} + \sum_{m \neq 0} c_m \sqrt{y} K_{r-\frac{1}{2}}(2\pi|m|y) \exp(2\pi imx),$$

and

$$T_n f(z) = by^r + b'y^{1-r} + \sum_{m \neq 0} b_m \sqrt{y} K_{r-\frac{1}{2}}(2\pi|m|y) \exp(2\pi imx).$$

Then $b = n^{\frac{1}{2}-r}\sigma_{2r-1}(n)c, \quad b' = n^{r-\frac{1}{2}}\sigma_{1-2r}(n)c'$, *and*

$$b_m = \sum_{\substack{d \mid \gcd(m,n) \\ d>0}} c_{\frac{mn}{d^2}}.$$

The sum is over all positive divisors d of the greatest common divisor of m and n.

(3) *The algebra generated by the Hecke operators is commutative and the* ***multiplication formula*** *is*

$$T_nT_m = \sum_{\substack{d \mid \gcd(m,n) \\ d>0}} T_{\frac{mn}{d^2}}.$$

(4) *Define the* ***Petersson inner product*** *for Maass cusp forms* $f, g \in \mathcal{SN}(SL(2,\mathbb{Z}), r(r-1))$ *by*

$$(f,g) = \int_{SL(2,\mathbb{Z})\backslash H} f(z)\overline{g(z)}\, y^{-2}dx\, dy.$$

Then $(T_nf,g) = (f,T_ng)$ *and thus the Hecke operators can be simultaneously diagonalized on the space* $\mathcal{N}(SL(2,\mathbb{Z}), r(r-1))$. *Note that in this case the Petersson inner product is the same as the usual* L^2*-inner product.*

(5) *Suppose that* $f \in \mathcal{N}(SL(2,\mathbb{Z}), r(r-1))$ *has the Fourier expansion*

$$f(z) = cy^r + c'y^{1-r} + \sum_{m\neq 0} c_m\sqrt{y}K_{r-\frac{1}{2}}(2\pi|m|y)\exp(2\pi imx)$$

and that f is an eigenfunction of all the Hecke operators; i.e.,

$$T_nf = u_nf, \quad \text{for } n = 1,2,3,\ldots.$$

Then

$$c_n = u_{|n|}c_{n/|n|} \text{ for } n \neq 0;$$

and the two associated Dirichlet series have ***Euler products***

$$L_f\left(s-\frac{1}{2}\right) = \sum_{m\neq 0} c_m|m|^{-s} = (c_1 + c_{-1}) \prod_{p \text{ prime}} (1 - u_pp^{-s} + p^{-2s})^{-1},$$

$$\begin{aligned} L_{f_x}\left(s+\frac{1}{2}\right) &= \sum_{m\neq 0} (2\pi im)c_m|m|^{-s-1} \\ &= 2\pi i(c_1 - c_{-1}) \prod_{p \text{ prime}} (1 - u_pp^{-s} + p^{-2s})^{-1}. \end{aligned}$$

Conversely, if $f \in \mathcal{N}(SL(2,\mathbb{Z}), r(r-1))$ *corresponds to Dirichlet series* L_f *and* L_{f_x} *with Euler products such as those above, then* f *is an eigenfunction for all the Hecke operators.*

Proof.

(1) First note that the Hecke operators clearly commute with the non-Euclidean Laplace operator. It is easy to see that $T_n f$ is invariant under $SL(2,\mathbb{Z})$ if f is invariant. For if M_n denotes the 2×2 integer matrices of determinant n, L runs through a full set of representatives for $SL(2,\mathbb{Z})\backslash M_n$, and $\gamma \in SL(2,\mathbb{Z})$, then $L\gamma$ also runs through a full set of representatives for $SL(2,\mathbb{Z})\backslash M_n$. Finally, it is clear that if $f(z)$ has at most polynomial growth in y, then so does $T_n f$, as y approaches infinity.

(2) Suppose that T_n is given by formula (3.119). First note that $\operatorname{Im}\left(\frac{az+b}{d}\right) = \frac{ny}{d^2}$ and $\operatorname{Re}\left(\frac{az+b}{d}\right) = \frac{ax+b}{d}$. It follows that

$$T_n f(z) = \frac{1}{\sqrt{n}} \sum_{\substack{n=ad, d>0 \\ b \pmod{d}}} \left(c(nyd^{-2})^r + c'(nyd^{-2})^{1-r}\right)$$
$$+ \frac{1}{\sqrt{n}} \sum_{\substack{n=ad, d>0 \\ b \pmod{d}}} \sum_{m \neq 0} c_m \frac{\sqrt{ny}}{d} K_{r-\frac{1}{2}}\left(\frac{2\pi|m|ny}{d^2}\right) e^{2\pi i m(ax+b)/d}.$$

The first sum is

$$y^r c n^{r-\frac{1}{2}} \sum_{0<d \mid n} d^{1-2r} + y^{1-r} c' n^{\frac{1}{2}-r} \sum_{0<d \mid n} d^{2r-1}$$
$$= c y^r n^{r-\frac{1}{2}} \sigma_{1-2r}(n) + c' y^{1-r} n^{\frac{1}{2}-r} \sigma_{2r-1}(n).$$

The second sum simplifies since the sum over b is zero unless d divides m. So the second sum is

$$\sqrt{y} \sum_{d \mid \gcd(m,n)} c_m K_{r-\frac{1}{2}}\left(\frac{2\pi|m|ny}{d^2}\right) e^{2\pi i m n x d^{-2}}.$$

Sum over $a = n/d$ instead of d and set $v = ma/d$ (so that $m = nva^{-2}$) to obtain

$$\sum_{\substack{0<a \mid \gcd(n,v) \\ v \in \mathbb{Z}}} c_{\frac{nv}{a^2}} \sqrt{y} K_{r-\frac{1}{2}}(2\pi|v|y) e^{2\pi i v x}.$$

This gives the stated result.

(3) The proof is easy if $\gcd(m,n) = 1$. For suppose the matrices

$$\begin{pmatrix} d_1 & d_{12} \\ 0 & d_2 \end{pmatrix}$$

run through a complete set of representatives for M_n modulo $SL(2,\mathbb{Z})$, as in formula (3.117) and that the matrices

$$\begin{pmatrix} c_1 & c_{12} \\ 0 & c_2 \end{pmatrix}$$

run through a complete set of representatives for M_m modulo $SL(2,\mathbb{Z})$. The product is

$$\begin{pmatrix} d_1c_1 & d_1c_{12}+c_2d_{12} \\ 0 & d_2c_2 \end{pmatrix}.$$

If d_{12} runs through a complete set of representatives $\bmod d_2$ and c_{12} runs through a complete set of representatives $\bmod c_2$, then $d_1c_{12}+c_2d_{12}$ runs through a complete set of representatives $\bmod d_2c_2$. Why?

To complete the proof of (3), we want to show that

$$T_{p^k}T_p = T_{p^{k+1}} + T_{p^{k-1}} \text{ for } k \geq 1. \tag{3.120}$$

Let us set up the following notation. We shall call two matrices **equivalent** and write $A \sim B$ if $f(Az) = f(Bz)$, for all $f \in \mathcal{N}(SL(2,\mathbb{Z}), r(r-1))$.[10]

In order to prove Eq. (3.120), we shall multiply matrices in the set of representatives for $SL(2,\mathbb{Z})\backslash M_n$ which are summed over in the definition (3.119) of the Hecke operator. We have, for $e+f=k$ two kinds of matrices to multiply. First we have:

$$\left.\begin{aligned} &\begin{pmatrix} 1 & a \bmod p \\ 0 & p \end{pmatrix}\begin{pmatrix} p^e & b \bmod p^f \\ 0 & p^f \end{pmatrix} \\ &= \begin{pmatrix} p^e & b \bmod p^f + p^f(a \bmod p) \\ 0 & p^{f+1} \end{pmatrix} = \begin{pmatrix} p^e & b \bmod p^{f+1} \\ 0 & p^{f+1} \end{pmatrix}. \end{aligned}\right\} \tag{3.121}$$

Also we obtain:

$$\left.\begin{aligned} &\begin{pmatrix} p & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} p^e & b \bmod p^f \\ 0 & p^f \end{pmatrix} \\ &= \begin{pmatrix} p^{e+1} & p(b \bmod p^f) \\ 0 & p^f \end{pmatrix} \sim \begin{pmatrix} p^e & b \bmod p^f \\ 0 & p^{f-1} \end{pmatrix} \\ &= \begin{pmatrix} p^e & b_1 \bmod p^{f-1} + p^{f-1}(c_1 \bmod p) \\ 0 & p^{f-1} \end{pmatrix} \\ &= \begin{pmatrix} 1 & c_1 \bmod p \\ 0 & 1 \end{pmatrix}\begin{pmatrix} p^e & b_1 \bmod p^{f-1} \\ 0 & p^{f-1} \end{pmatrix} \\ &\sim \begin{pmatrix} p^e & b_1 \bmod p^{f-1} \\ 0 & p^{f-1} \end{pmatrix}. \end{aligned}\right\} \tag{3.122}$$

---

[10] This amounts to saying that $c\gamma A = B$ for some $\gamma \in SL(2,\mathbb{Z})$, $c > 0$.

Equation (3.121) leads to operators in $T_{p^{k+1}}$, except that the $f+1=0$ term is missing. Equation (3.122) gives that term when $f=1$. And what remains in Eq. (3.122) gives operators from $T_{p^{k-1}}$, each of them taken p times, when we take account of the $c_1 \bmod p$. This completes the proof of formula (3.120). Why?

Mathematical induction and Eq. (3.120) imply that

$$T_{p^e} T_{p^f} = \sum_{t=0}^{\min(e,f)} T_{p^{e+f-2t}}. \tag{3.123}$$

From this result, part (3) follows easily.

Note also that Eq. (3.120) implies the following formula for the formal power series in the indeterminate X (for prime p):

$$\sum_{e \geq 0} T_{p^e} X^e = (I - T_p X + X^2)^{-1}. \tag{3.124}$$

To prove this, you must simply multiply both sides by $(I - T_p X + X^2)$ and make use of Eq. (3.120). Formula (3.124) will lead to Euler products for L-functions corresponding to eigenfunctions of all the Hecke operators.

Exercise 3.6.13. Fill in the details that were omitted in the proof of part (3); e.g., in the proofs of Eqs. (3.123) and (3.124).

(4) This argument goes back to Petersson [511]. Recall that M_n is the set of 2×2 integral matrices of determinant n and $\Gamma = SL(2,\mathbb{Z})$. Thus Eq. (3.119) says that the Hecke operator T_n is a sum over $\Gamma \backslash M_n$ and thus

$$n^{1/2}(T_n f, g) = \sum_{A \in \Gamma \backslash M_n} \int_{\Gamma \backslash H} f(Az)\overline{g(z)} y^{-2} dx\, dy.$$

For $A \in M_n$, set $h(z) = f(Az)$. Note that $h(Bz) = h(z)$, for all B in the congruence subgroup

$$\Gamma(n) = \{B \in SL(2,\mathbb{Z}) = \Gamma \mid B \equiv I (\mathrm{mod}\, n)\}$$

(see Exercise 3.6.14).

Since a fundamental domain $\Gamma(n)\backslash H$ can be taken to be $[\Gamma : \Gamma(n)]$ copies of $\Gamma \backslash H$ (by Exercise 3.3.9 of Sect. 3.3), we have

$$\begin{aligned}
\sqrt{n}(T_n f, g) &= \sum_{A \in \Gamma \backslash M_n} \frac{1}{[\Gamma : \Gamma(n)]} \int_{\Gamma(n)\backslash H} f(Az)\overline{g(z)} \frac{dx\, dy}{y^2} \\
&= \sum_{A \in \Gamma \backslash M_n} \frac{1}{[\Gamma : A\Gamma(n)A^{-1}]} \int_{A\Gamma(n)A^{-1}\backslash H} f(w)\, \overline{g(A^{-1}w)}\, \frac{du\, dv}{v^2},
\end{aligned}$$

setting $w = u + iv = Az$ and noting that $[\Gamma : A\Gamma(n)A^{-1}] = [\Gamma : \Gamma(n)]$. Part (b) of Exercise 3.6.14 tells us that the last quantity is none other than $(f, T_n g)\sqrt{n}$ and part (4) of Theorem 3.6.4 is proved.

Exercise 3.6.14.

(a) Suppose that A is a 2×2 integral matrix of determinant n and that f is a modular form for $\Gamma = SL(2,\mathbb{Z})$. Define $h(z) = f(Az)$. Prove that $h(z)$ is a modular form for the congruence subgroup $\Gamma(n)$.

(b) If A runs through a complete set of representatives for $\Gamma\backslash M_n$, where M_n is the set of integral 2×2 matrices of determinant n, $\Gamma = SL(2,\mathbb{Z})$, show that nA^{-1} also runs through a complete set of representatives for M_n/Γ. Then show that there is a common set of representatives for M_n/Γ and $\Gamma\backslash M_n$.

Hint.

(a) Note that $nA^{-1} \in \mathbb{Z}^{2\times 2}$ and that $B \in \Gamma(n)$ implies that $nABA^{-1} \equiv 0 (\operatorname{mod} n)$. But then $ABA^{-1} \in SL(2,\mathbb{Z})$.

(b) For the last part of this exercise see Shimura [589, pp. 53–54]. You need to know that the number of left Γ-cosets of $\Gamma A\Gamma$ is the same as the number of right $\Gamma-$cosets of $\Gamma A \Gamma$. It is easy to prove that these numbers are the same in the case at hand.

(5) First suppose that $T_n f = u_n f$. From part (2) of the theorem, it follows that

$$c_n = u_n c_1 \quad \text{and} \quad c_{-n} = u_n c_{-1}.$$

Thus

$$L_f\left(s - \frac{1}{2}\right) = \sum_{n\neq 0} c_n |n|^{-s} = (c_1 + c_{-1}) \sum_{n\geq 1} u_n n^{-s}.$$

and there is an analogous formula for L_{f_x}. The Euler product follows easily from formula (3.124).

Conversely, suppose that L_f and L_{f_x} have Euler products of the indicated type; e.g.,

$$L_f\left(s - \frac{1}{2}\right) = \sum_{m\neq 0} c_m |m|^{-s} = (c_1 + c_{-1}) \prod_{p \text{ prime}} (1 - u_p p^{-s} + p^{-2s})^{-1}.$$

Then the following Dirichlet series has no terms $|m|^{-s}$ such that p divides m:

$$(1 - u_p p^{-s} + p^{-2s}) \sum_{m\neq 0} c_m |m|^{-s} = \sum_{m\neq 0} (c_m - u_p c_{m/p} + c_{m/p^2}) |m|^{-s}.$$

Plug in $m = np$ to see that $c_{np} - u_p c_n + c_{n/p} = 0$. Thus

$$c_{np} + c_{n/p} = u_p c_n.$$

The left-hand side of this last equality is the nth Fourier coefficient of $T_p f$ and the right-hand side is the nth Fourier coefficient of $u_p f$. Thus

$$T_p f = u_p f \quad \text{for all primes } p.$$

By formula (3.120), we know that $T_{p^e} = T_p T_{p^{e-1}} - T_{p^{e-2}}$, if $e \geq 2$. Thus we can show that f is an eigenfunction for T_{p^e} by mathematical induction. Then $T_m T_n = T_{mn}$ for $\gcd(m,n) = 1$ completes the proof of the converse in part (5), and the proof of the entire theorem.

□

Exercise 3.6.15. Show directly that both the holomorphic and the nonholomorphic Eisenstein series $G_k(z)$ and $E_s(z)$ are eigenfunctions of the appropriate Hecke operators T_n.

Hint. *Since the Hecke operator is a sum over* $\Gamma \backslash M_n$ *and the Eisenstein series is a sum over* $\Gamma_\infty \backslash \Gamma$*, the problem involves the determination of representatives of the quotient* $\Gamma_\infty \backslash M_n$*. That is, you must sum over* a, c *with* $\gcd(a,c) = t$*, for* t *dividing* n*. Then you need to notice that the solutions* b, d *of* $ad - bc = n$ *are not unique. Elementary number theory says there are* t *of them modulo* Γ_∞*. For set* $a = a_1 t,\ c = c_1 t,\ b = b_0 + a_1 u,\ d = d_0 + c_1 u,\ u \in \mathbb{Z}$*. Then*

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} a & b_0 \\ c & d_0 \end{pmatrix} \begin{pmatrix} 1 & u/t \\ 0 & 1 \end{pmatrix}.$$

You can check your calculation by noting that the eigenvalues are essentially the coefficients of the Fourier expansions of $G_k(z)$ *and* $E_s(z)$*.*

We have given a very elementary discussion of Hecke operators. It is possible to give an abstract definition of the Hecke algebra associated to a group G and subgroup H, assuming that each H-double coset is a finite union of right H-cosets. See Krieg [373] and Shimura [589]. This allows one to do Hecke theory for congruence subgroups, quaternion groups, and higher rank groups like $SL(n, \mathbb{Z})$.

3.6.6 Rankin–Selberg Method, Distribution of Horocycles in the Fundamental Domain, and Modular Symbols

In this section we give sketchy discussions of various topics that have been important for work on modular forms—the Rankin–Selberg method, the uniform distribution of horocycles in the fundamental domain, and modular symbols.

Example 3.6.5. ***(The Rankin–Selberg Method)*** There is another sort of Dirichlet series that can be associated to a modular form. We used this idea already in Exercise 3.5.8 of Sect. 3.5. The method goes back to Rankin [534] and Selberg [572], who used the method to obtain good estimates (though not so good as the Ramanujan–Petersson conjecture) for the Fourier coefficients a_n of holomorphic cusp forms $f(z)$ of weight k,

$$f(z) = \sum_{n\geq 1} a_n \exp(2\pi i n z),$$

by considering

$$L_{f^*}(s) = \sum_{n\geq 1} |a_n|^2 n^{1-k-s}. \tag{3.125}$$

The trick is to note that

$$(4\pi)^{1-s-k}\Gamma(s+k-1)L_{f^*}(s) = \int_{0<y,|x|\leq\frac{1}{2}} y^{k+s}|f(z)|^2 y^{-2}\,dy\,dx. \tag{3.126}$$

The integral here is over a fundamental domain for $\Gamma_\infty\backslash H$, where Γ_∞ is the subgroup of $\Gamma = SL(2,\mathbb{Z})$ fixing $i\infty$, defined by formula (3.82).

Next note that we can write the right-hand side of Eq. (3.126) as

$$\int_{\Gamma\backslash H} \sum_{\gamma\in\Gamma_\infty\backslash\Gamma} \operatorname{Im}(\gamma z)^s y^k\,|f(z)|^2\,y^{-2}\,dx\,dy = \int_{\Gamma\backslash H} E_s(z) y^k\,|f(z)|^2\,y^{-2}\,dx\,dy. \tag{3.127}$$

Exercise 3.6.16. Prove formulas (3.126) and (3.127).

Then the analytic continuation of the Eisenstein series leads to the desired estimates for the coefficients a_n, using methods from analytic number theory. This method works for congruence subgroups. However, Selberg [570] gives examples of discrete groups Γ' such that the Eisenstein series has poles in $(1-\delta,1)$. In this same paper, Selberg shows how to get around this problem for Γ' of finite index in Γ, by going from scalar modular forms for Γ' to vector-valued modular forms for Γ.

Part of Deligne's proof of the Weil conjectures was motivated by the Rankin–Selberg argument (see Katz [344]). Many other applications of the Rankin–Selberg method have appeared. Some examples are Andrianov [8], Moreno [477], Novodvorsky and Piatetskii-Shapiro [496], Piatetskii-Shapiro [515], Shimura [589–591], Stark [626, Part II], and Zagier [750]. See also Iwaniec [319].

There are other classical methods of estimating Fourier coefficients of modular forms using Poincaré series and Kloosterman sums (see Petersson [512] and Selberg [570]).

Iwaniec (in joint work with Sarnak) obtains a bound on $\|u\|_\infty < C\lambda^{7/32}$ for cuspidal Maass waveforms u with $\Delta u = \lambda u$ and which are eigenfunctions of all the Hecke operators. The standard bound is $\lambda^{\frac{1}{4}}$. See Iwaniec [318].

R.A. Smith [608] showed that if f is a cuspidal Maass waveform in the space $\mathcal{SN}(SL(2,\mathbb{Z}),r(r-1))$, $r=\frac{1}{2}+it$, $t>0$, and f is a common eigenfunction for all the Hecke operators, normalized so that its first Fourier coefficient c_1 has absolute value 1, then the L^2-norm of f on the fundamental domain satisfies the inequality

$$\frac{1}{6}\,|\Gamma(it)| \leq (f,f)^{\frac{1}{2}} \leq 2\,|\Gamma(1+it)| \quad \text{for } t \geq 200.$$

Thus the norm of f decreases exponentially as $t \to \infty$.

The following exercise concerns the horocycle $\{x+iy \mid x \in \mathbb{R}\}$, for fixed $y>0$. It proves this horocycle to be equidistributed in the fundamental domain for the modular group as $y \to 0+$. Strömbergsson [646] improves on this result.

Exercise 3.6.17 (Uniform Distribution of Horocycles in the Fundamental Domain as the Horocycle Approaches the Real Axis). Let $C_y \subset \Gamma\backslash H$ be the curve in the fundamental domain for Γ obtained from the horizontal line at height y; i.e., the horocycle $\Gamma_\infty\backslash\mathbb{R}+iy$. Then C_y is a closed curve in the fundamental domain having Poincaré length $1/y$. Show that C_y fills up $\Gamma\backslash H$ in a very uniform way as $y \to 0^+$, in the sense that, for any open set U in $\Gamma\backslash H$, we have

$$\frac{\text{length}(C_y \cap U)}{1\text{ength}(C_y)} \sim \frac{\text{area}(U)}{\text{area}(\Gamma\backslash H)} \quad \text{as } y \to 0+\,. \tag{3.128}$$

Here all lengths and areas are non-Euclidean!

Hints (see Zagier [750]). Use $k=0$ and $f(z)=\chi_U(z)=$ the indicator function of U, which is defined to be 1 on U and 0 outside U, in formulas (3.126) and (3.127). Then

$$C(U,y)=\int_0^1 \chi_U(x+iy)dx=\frac{\text{length}(C_y \cap U)}{\text{length}(C_y)}$$

and

$$I(U,s)=\int_{y=0}^{\infty} C(U,y)y^{s-2}dy=\int_{\Gamma\backslash H} \chi_U(z)E_s(z)y^{-2}dx\,dy.$$

Use a Tauberian theorem for Mellin transforms, after noting that

$$\text{Res}_{s=1}I(U,s)=\frac{\text{area}(U)}{\text{area}(\Gamma\backslash H)}.$$

Zagier [750] goes on to show that if the error term in Eq. (3.128) is $O(y^{3/4-\varepsilon})$ the Riemann hypothesis is true.[11]

---

[11] The notation "$O(y^p)$" stands for a function which when divided by y^p is bounded as $y \to \infty$.

One can view the integral

$$\int_0^{i\infty} f(z)z^s dz$$

as a special case of the period integral (for $s \in \mathbb{Z}$)

$$\int_a^b f(z)z^s dz$$

with $a, b \in P_\Gamma = \{\text{cusps of the discrete group } \Gamma\}$. This observation allows Manin [448–451] to develop a systematic way of calculating these integrals or **modular symbols**, making use of facts about cohomology, Farey fractions, and continued fractions. Write $\{a,b\}_\Gamma$ for the modular symbol or the homology class determined by the geodesic joining the cusps a and b in P_Γ. Then one has

$$\{\gamma a, \gamma b\}_\Gamma = \{a,b\}_\Gamma \quad \text{for all } \gamma \in \Gamma,$$

and

$$\{a_1, a_2\}_\Gamma + \cdots + \{a_k, a_1\}_\Gamma = 0 \quad \text{for all sequences of cusps } a_1, \ldots, a_k.$$

The second sort of relation brings in the Farey fractions. One can also transfer the action of the Hecke operators to the homology. As a result (after some modification of the theory when the weight of the cusp form is not 2), one can prove, for example, the following fact about ratios of L-series corresponding to the cusp form $\Delta(z)$ of weight 12 (the discriminant defined in Exercise 3.6.4 of Sect. 3.4).

$$\left(\int_0^{i\infty} \Delta(z)dz : \int_0^{i\infty} \Delta(z)z^2dz : \int_0^{i\infty} \Delta(z)z^4dz\right) = \left(1 : \frac{-691}{2^2 3^4 5} : \frac{-691}{2^3 3^2 5 \cdot 7}\right).$$

Such work was begun by Eichler [153] and Shimura [592]. A reference is Lang [391]. It is proved, for example, that ratios of certain L-values in the critical strip lie in the field generated by the Fourier coefficients of the corresponding modular form. This fits into a general philosophy of Deligne (see Deligne's article in Borel and Casselman [52, Vol. 2, pp. 313–346] and Zagier [752, pp. 118–120]). Another paper on this subject is Razar [538].

Voronin [709] considers a similar sort of Dirichlet series to that in formula (3.105). Suppose that $f(z) = \Delta(z)$, as in formula (3.68). Let C be the vertical line in H starting at the point a on the real axis. Thus C is a geodesic. Consider the integral $I(s) = \int_C (-i(z-a))^{s-1} f(z)dz$. It is easy to see that for Re $s > 12$, $I(s) = i(2\pi)^{-s}\Gamma(s)\Phi(s)$, where

$$\Phi(s) = \sum_{n \geq 1} \tau(n) \exp(2\pi i n a) n^{-s}, \text{ Re } s > 12.$$

Here $\tau(n)$ denotes the Ramanujan tau-function defined by formula (3.66). Voronin manages to obtain an analytic continuation of $\Phi(s)$ as a meromorphic function in the whole complex s-plane.

Exercise 3.6.18 (Fourier Expansions of Other Eisenstein Series). Compute the Fourier expansions for some other Eisenstein series; e.g., that for the leaky torus in Exercise 3.6.14 of Sect. 3.3 or for some congruence subgroup.

If you have access to a computer, the following exercise should be possible. It illuminates what is happening in Exercise 3.6.17 above, as well as in Fig. 3.39.

Exercise 3.6.19 (H. Stark) (Computer-Generated Movie of a Non-Euclidean Shock Wave). Consider the following points on a horocycle: $z_j(y) = iy + j/N,\ j = 1, 2, \ldots, N$, holding N fixed. Find $\gamma \in SL(2,\mathbb{Z})$ such that $\gamma z_j(y) = w_j(y)$ lies in the standard fundamental domain for $SL(2,\mathbb{Z})\backslash H$ (as in Fig. 3.14). Start with $y = 2$. Let y approach zero from above and watch what happens to the points $w_1(y), \ldots, w_N(y)$. At first, you see points on a horizontal line segment of length 1 and height 2. As the segment moves down ($y \to 0+$), the individual points $z_j(y)$ move along geodesics. You can consider the points as forming a non-Euclidean shock wave. The wave reflects from the boundaries of the fundamental domain as y passes through $y = 1$. Then for smaller values of y, more reflections occur and the picture begins to look quite chaotic. The maximum amount of chaos appears to occur around $y = 1/N$. After that the picture begins to become less random. Ultimately the points form various horizontal line segments which move up to the cusp at infinity. For example, if $N = 51$, there are four line segments corresponding to points with reduced denominators $1, 3, 17, 51$. This is related to the rational cusps approached by the $z_j(y)$, as $y \to 0$. This movie can be generated on your favorite computer. Stark created this movie on his Atari ST which is still operating.

3.7 Harmonic Analysis on the Fundamental Domain

> The Selberg theory allows us to actually perform calculations on noncocompact but finite volume universes which are of interest to those working in general relativity.
>
> —From Hurt [308, p. xiv].

3.7.1 *Introduction*

In this section we seek to describe the non-Euclidean analogue of Fourier series for $L^2(SL(2,\mathbb{Z})\backslash H)$ as well as two related analogues of the Poisson summation formula. See Sects. 1.3 and 1.4 for the Euclidean versions and compare them with the non-Euclidean results given in Theorems 3.7.1, 3.7.3, and 3.7.4 which follow. We shall discuss various applications in number theory, analysis, and geometry.

Number-theoretic applications include a non-Euclidean analogue of the circle problem on the asymptotics of the number of lattice points inside a circle of radius x, as x approaches infinity (cf. Theorem 1.3.5 of Sect. 1.3 and formula (3.136) that follows). We will also consider the asymptotics of units in real quadratic fields (in Theorem 3.7.6). There are other number-theoretic applications that will not be discussed here; e.g., the Eichler–Selberg trace formula for the trace of a Hecke operator, the Riemann hypothesis for Selberg's zeta function, formulas for dimensions of spaces of holomorphic modular forms, and Langlands' work on the Artin conjecture for Artin L-functions using twisted trace formulas. More analytic applications include the asymptotics of the eigenvalues of the Laplacian on $L^2(SL(2,\mathbb{Z})\backslash H)$, which is another non-Euclidean version of Theorem 1.3.5 of Sect. 1.3 (see Theorem 3.7.5 which follows), and the solution of the non-Euclidean heat equation on $SL(2,\mathbb{Z})\backslash H$ (Exercise 3.7.16). The Selberg trace formula (Theorem 3.7.4) provides a duality between the eigenvalues of the Laplacian on $M = \Gamma\backslash H$ and closed geodesics in M (if the contribution from the elliptic and parabolic elements of Γ is small). Thus one obtains geometric information about M.

The last remarks are related to the question: "**Can one hear the shape of a drum**?" asked by Kac in [332] and discussed in Sect. 1.3. As we saw in that section, one can consider the problem for a plane membrane M, held fixed along its boundary C. If the membrane is set in motion and $u(z,t)$ denotes the vertical displacement above $z \in M$ at time t, then u satisfies the wave equation

$$\Delta_z u(z,t) = u_{tt}(z,t), \quad \text{for } z \text{ in } M,$$

$$u(z,t) = 0, \text{ for } z \text{ on the boundary } C.$$

The solutions will be superpositions of normal modes $\exp(i\omega_n t)v_n(z)$. Here $v_n(z)$, $n = 1,2,3,\ldots$, denotes a complete orthonormal set of eigenfunctions of the Laplacian on M which vanish on the boundary C, with eigenvalue $-\omega_n^2 = \lambda_n$, where the spectrum of Δ is $\{\lambda_n,\ n = 1,2,3,\ldots\}$. This comes from Theorem 1.3.7 of Sect. 1.3 on the spectra of compact self-adjoint operators. One hears the normal modes or pure tones ω_n corresponding to the eigenvalues of the Laplacian.

The question is then, "What can you say about the geometry of M if you know the spectrum λ_n, $n = 1,2,3,\ldots$?" It is, of course, possible to pose the analogous question for any Riemannian manifold. Milnor showed that there are two noncongruent lattices L in $\mathbb{R}^{16}$ such that the corresponding tori $\mathbb{R}^{16}/L$ have the same spectra. Thus you cannot hear everything about a torus. Marie-France Vignéras [701] found analogous examples $\Gamma\backslash H$, with Γs coming from quaternion algebras. In 1985, Sunada found a method which uses group representations to construct examples of drums that cannot be heard (see [649]). Buser [74] found such examples of isospectral manifolds coming out of graph-theoretic analogues, which themselves originate in examples of nonisomorphic algebraic number fields with the same Dedekind zeta function. But these Buser manifolds were not flat and so did not answer the Kac question. In 1992, Carolyn Gordon et al. [227] showed how to flatten the Buser examples and thus that there are (nonconvex) planar drums with shapes that cannot be heard. See my books [668, 671] for more information on the graph theory analogues.

Although you cannot expect to hear everything about M from the eigenvalues of the Laplacian on M, there are still many geometric quantities that can be determined from the spectrum. For example, Weyl showed (in the Euclidean case) that you can hear the area of M. A non-Euclidean analogue of this result is contained in Theorem 3.7.5 which follows.

One can also consider the problem of material diffusing through a plane region M with boundary C, starting at a position z, in such a way that it is absorbed upon hitting the boundary curve C. The density of matter $p_M(z,w;t)$ at position w will satisfy the **heat or diffusion equation**

$$\begin{aligned}\frac{\partial}{\partial t}p_M(z,w;t) &= \Delta_w p_M(z,w;t),\\ p_M(z,w;t) &\to 0 \quad \text{as } w \text{ approaches the boundary},\\ p_M(z,w;t) &\to \delta(z-w) \quad \text{as } t \to 0+\,.\end{aligned}$$

Here $\delta(x)$ denotes the Dirac delta distribution centered at the origin in the plane. If v_n and λ_n are as above, then one can express p_M as follows:

$$p_M(z,w;t) = \sum_{n\geq 1} \exp(\lambda_n t)\overline{v_n(z)}v_n(w).$$

As $t \to 0+$, it is reasonable to expect that "particles of the diffusing stuff will not have had enough time to have felt the influence of the boundary C. As particles begin to diffuse they may not be aware, so to speak, of the disaster that awaits them when they reach the boundary," according to Kac [330, p. 481]. Thus one expects that

$$p_M(z,w;\ t) \sim \frac{1}{4\pi t}\exp\left[-\frac{\|z-w\|^2}{4t}\right] \quad \text{as } t \to 0+$$

(see Exercise 1.2.13 of Sect. 1.2). It will follow upon integration that

$$\sum_{n\geq 1}\exp(-\lambda_n t) \sim \text{area }(M)/4\pi t, \qquad \text{as } t \to 0+\,.$$

This is the Laplace transform of the result of Weyl mentioned earlier, which allows one to hear the area of the drum. We consider a non-Euclidean analogue of this result in the proof of Theorem 3.7.5 below. There has been much work on further terms in the asymptotic expansion given above (see Berger [36] and Molchanov [473] for references). Kac [loc. cit.] provides connections between the asymptotic result above and an asymptotic formula relating quantum-statistical mechanics and classical mechanics, as well as connections with the Wiener integral. This latter integral interprets diffusion in terms of a measure defined on the space of all continuous curves emanating from the origin. The measure is defined to agree with the measure coming from the Einstein–Smoluchowski theory of Brownian motion.

These considerations have many connections with physics, as Molchanov [473, p. 7] describes in the following quotation: "In the physics literature the idea of 'integrating along trajectories' as a method of studying spectra has found broad application in quantum mechanics, mainly in the work of the Feynman school; the monograph (Feynman [178]) is devoted to the theory of Feynman integrals. It concerns 'measures' in a space of trajectories, which are constructed from a fundamental solution of the Schrödinger wave equation. A mathematically rigorous foundation for the Feynman theory, at least in certain of its facets, is even now still wanting (see the discussion in [Daletskii [116])." Another reference on Feynman integrals is the paper of DeWitt [131]. Hurt [308, 309], Gutzwiller [239–241], Dowker [138, 139] give some insight into the connections between the Selberg trace formula and the Feynman integral picture, and more.

Most of the results on harmonic analysis on $\Gamma\backslash H$, for Fuchsian groups Γ of the first kind, are due to Atle Selberg, who lectured on them at Göttingen in 1954. The manuscript of the part of the lectures concerning compact fundamental domains appeared to have been lost for many years thanks to the absence of Xerox machines. However, the second part of the lectures concerning noncompact fundamental domains such as that for $SL(2,\mathbb{Z})$ was to be found in the Göttingen library (this is Selberg [573]) and now is in Selberg's collected works [574, Vol. I, pp. 626–674] along with Selberg's interesting commentary on the lectures. Many more things are in the Selberg archive section of the Institute for Advanced Study website

http://publications.ias.edu/selberg.

W. Roelcke was working independently on the subject around the same time (see Roelcke [545]). Previously Delsarte worked out some of the theory for the compact fundamental domains (see Delsarte [129, Tome II, pp. 599–601, 829–845]), motivated by non-Euclidean lattice point problems.

In this section (and indeed this chapter), we have chosen to consider only the example of $\Gamma = SL(2,\mathbb{Z})$. In one respect, this is a nice example, for one can make everything in the trace formula very concrete. However, in another respect, this example, is a difficult one, for it produces parabolic terms in the trace formula—parabolic terms that are rather complicated. Even the elliptic terms are not trivial. So the reader might want to replace Γ by some nice arithmetic group with compact fundamental domain and no elliptic fixed points. Examples of such groups were discussed in Sect. 3.3.

General references for this section include Borel and Casselman [52], Borel and Mostow [54], Bump's introduction to the trace formula in Bernstein and Gelbart [37, pp. 153–196], Bump [71], Chavel [84], Duistermaat et al. [143], Elstrodt [155, 156], Elstrodt et al. [157–160], Faddeev [171], Fay [176], Gangolli [191], Gangolli and Warner [195, 196], Gelfand et al. [203], Godement [213–216], Goldfeld and Hundley [219], Goldfeld and Husemoller [220], Hejhal [261–271], Hurt [308], Iwaniec [318, 319], Kubota [375], Lang [389], Lapid [395], Lax and Phillips [400], Marklof's introduction to Selberg's trace formula in Bolte and Steiner [47, pp. 83–120], McKean [461], Roelcke [544–546], Sarnak [555–557], Selberg [569, 571, 573, 574], Subia [647], Tamagawa [656], Venkov [693, 695], Venkov et al. [696], Marie-France Vignéras [700], Voros [710], Dorothy Wallace [711], and Warner [718]. The Sarnak homepage at IAS also has enough ideas to keep you busy for a very long time. And don't forget the Langlands archives at IAS.

3.7.2 Spectral Resolution of Δ *on* $SL(2,\mathbb{Z})\backslash H$

We shall begin by discussing the spectral resolution of the non-Euclidean Laplacian on the fundamental domain. This consists of an expansion of an "arbitrary" function $f: SL(2,\mathbb{Z})\backslash H \to \mathbb{C}$ in eigenfunctions of Δ. Such an expansion is a non-Euclidean analogue of Fourier series, but it involves a mixture of series and integrals. Because the fundamental domain is not compact, the eigenvalue problem for Δ on $SL(2,\mathbb{Z})\backslash H$ is said to be **singular**. There may be continuous or discrete spectra (or both) in such cases. For example, recall the eigenvalue problem arising from the quantum mechanics of the hydrogen atom (see formula (2.14) of Sect. 2.1). See Stakgold [618, 619] and Titchmarsh [679] for other examples of eigenvalue problems with mixtures of continuous and discrete spectra.

The reader should recall from Sect. 3.5 the basic facts about the spectrum of the non-Euclidean Laplacian on $SL(2,\mathbb{Z})\backslash H$. Let us now collect the basic ingredients for our cake.

The Various Parts of the Spectrum of the Non-Euclidean Laplacian on $SL(2,\mathbb{Z})\backslash H$:

(a) The **discrete spectrum of constants and cusp forms**, with orthonormal basis

$$\left.\begin{array}{l} \{v_n\}_{n\geq 0},\ v_0=\sqrt{\frac{3}{\pi}},\ v_n = \quad \text{cusp form},\ n\geq 1. \\ \Delta v_n = s_n(s_n-1)v_n,\ s_n\in \frac{1}{2}+i\mathbb{R}, \\ v_n(z)\sim 0,\ \text{as } z\to\infty,\ \text{for } n\geq 1. \end{array}\right\} \tag{3.129}$$

(b) The **continuous spectrum of Eisenstein series,** with the Eisenstein series defined in formula (3.81),

$$\left.\begin{array}{l} \{E_s \mid s=\frac{1}{2}+it,\ t\in\mathbb{R}\}, \\ \Delta E_s = s(s-1)E_s, \\ E_s(z)\sim y^s+\frac{\Lambda(1-s)}{\Lambda(s)}y^{1-s}, \quad \text{as } z\to\infty, \\ \quad \text{where}\ \ \Lambda(s)=\pi^{-s}\Gamma(s)\zeta(2s). \end{array}\right\} \tag{3.130}$$

Exercise 3.7.1. Show that the three spaces of Eisenstein series, constants, and cusp forms, respectively, are mutually orthonormal.

Note. Throughout this section the inner product on $L^2(SL(2,\mathbb{Z})\backslash H)$ is

$$(f,g)=\int_{SL(2,\mathbb{Z})\backslash H} f(z)\overline{g(z)}y^{-2}dx\,dy. \tag{3.131}$$

The Eisenstein series are not in $L^2(SL(2,\mathbb{Z})\backslash H)$ by Exercise 3.5.11 of Sect. 3.5, but they are in $L^1(SL(2,\mathbb{Z})\backslash H)$, when s is the critical strip. See also Exercise 3.5.12 of Sect. 3.5 and Selberg [571, pp. 183–184] and [574, Vol. I, pp. 650 and 672] for a way of truncating Eisenstein series in order to make sense of inner product formulas involving them.

Our first goal is to prove the following theorem.

Theorem 3.7.1 (The Roelcke–Selberg Spectral Resolution of the Laplacian Δ on L^2 (SL(2, $\mathbb{Z}$)$\backslash H$)). *Any* f *in* $L^2(SL(2,\mathbb{Z})\backslash H)$ *has the non-Euclidean Fourier expansion*

$$f(z) = \sum_{n\geq 0}(f,v_n)v_n(z) + \frac{1}{4\pi i}\int_{\mathrm{Re}\,s=\frac{1}{2}}(f,E_s)E_s(z)ds.$$

Here we use the notation of Eqs. (3.129), (3.130), and (3.131).

Before proving Theorem 3.7.1, note that the spectral measure $(4\pi i)^{-1}ds$ for the continuous part of the decomposition is computed from the asymptotics and functional equation of the Eisenstein series in accordance with the principle that worked in Theorems 3.2.1 and 3.2.2 of Sect. 3.2. For we know from Exercise 3.5.2 of Sect. 3.5 that

$$E_s(z) \sim y^s, \quad \text{as } y \to \infty, \quad \text{for } s \text{ fixed with } \mathrm{Re}\,s > 1.$$

And formula (3.68) shows that $E_s(z)$ has a functional equation relating s to $1-s$. Thus we find that the spectral measure should be

$$\frac{\text{spectral measure for usual Mellin inversion in Sect. 1.4}}{\text{number of functional equations}} = \frac{1}{4\pi i}.$$

Lemma 3.7.1, which follows, gives a rigorous derivation of this spectral measure. Since $E_s(z)$ has a pole at $s=1$, there is more happening in Theorem 3.7.1 than we saw in Theorems 3.2.1 and 3.2.2 of Sect. 3.2. And we will find that Cauchy's residue theorem throws $\mathrm{Res}_{s=1}E_s(z)$ into the discrete part of the spectrum of Δ on $SL(2,\mathbb{Z})\backslash H$.

The proof that follows is decomposed into a sequence of lemmas that will only end with Theorem 3.7.2 concerning the discrete part of the spectrum. The method is that of R. Godement (see Borel and Mostow [54, pp. 211–234]). Other references for this discussion are Kubota [375, Chap. 5, including footnotes and a remark after Theorem 7.5.5] and Lang [389, Chap. 13].

The main idea of the proof is to reduce to the ordinary Mellin inversion formula by making use of a series we shall call the "incomplete theta series."

Definition 3.7.1. The **incomplete theta series** or **Poincaré series** $T\psi(z)$ for a sufficiently nice function $\psi:\mathbb{R}^+\to\mathbb{C}$ is given by

$$T\psi(z) = \sum_{\gamma\in\Gamma_\infty\backslash\Gamma}\psi(\mathrm{Im}\,(\gamma z)) \quad \text{for} \quad z\in H. \tag{3.132}$$

Here $\Gamma_\infty = \left\{\pm\begin{pmatrix}1 & n\\ 0 & 1\end{pmatrix}\middle|\, n\in\mathbb{Z}\right\}$ as in formula (3.82).

The following exercise gives some explanation of the word "nice" in the preceding definition.

Exercise 3.7.2.

(a) Show that if ψ has compact support, then the series (3.132) has only finitely many terms.
(b) Show that if there exists a $k > 1$ such that $|\psi(y)| \leq y^k$, for all $y > 0$, then series (3.132) converges.
(c) Give conditions on ψ that suffice to put $T\psi$ in $L^2(SL(2,\mathbb{Z})\backslash H)$.

Exercise 3.7.3 (An Adjoint for the Operator T).

(a) Define the **constant coefficient operator** on $f \in L^2(SL(2,\mathbb{Z})\backslash H)$ by

$$Cf(y) = \int_{-\frac{1}{2}}^{\frac{1}{2}} f(x+iy)dx.$$

Show that this operator pulls out the constant term in the Fourier expansion

$$f(z) = \sum_{n=-\infty}^{+\infty} c_n(y)e^{2\pi inx}.$$

That is, show that $(Cf)(y) = c_0(y)$.

(b) Show that if T denotes the incomplete theta series (3.132), and the inner product is as in formula (3.131)

$$(T\psi, f) = \int_{y>0} \psi(y)\overline{(Cf)(y)}\, y^{-2}dy,$$

assuming f is real-valued.

Hint. *Start with*

$$\int_{\Gamma\backslash H} \sum_{\Gamma_\infty\backslash\Gamma} = \int_{\Gamma_\infty\backslash H}.$$

This same idea was used in formulas (3.126) and (3.127) during the discussion of the Rankin–Selberg method.

Lang [389, pp. 241–243] notes that the adjointness relation in the last exercise happens for any pair of closed subgroups Γ, N of a Lie group (or just a locally compact topological group) G. Here

$$N = \left\{ \pm \begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix} \middle| \, x \in \mathbb{R} \right\}, \ \Gamma = SL(2,\mathbb{Z}), \ G = SL(2,\mathbb{R}).$$

And $N\backslash G \cong \mathbb{R}^2 - 0$ via the identification $Ng \to (0,1)g, \ g \in G$. The G-invariant measure on $N\backslash G$ is Lebesgue measure on $\mathbb{R}^2$. Lang shows that the same calculations that give part (b) of Exercise 3.7.3 say that

$$(T\psi, f)_{\Gamma\backslash G} = \int_{\Gamma\backslash G} T\psi(y)\overline{f(y)}dy = \int_{N\backslash G} \psi(y) \int_{\Gamma\cap N\backslash N} \overline{f(ny)}dn\, dy$$
$$= (\psi, Cf)_{N\backslash G},$$

defining

$$T\psi(y) = \int_{\gamma\in N\cap\Gamma\backslash\Gamma} \psi(\gamma y)d\gamma \quad \text{and}\, Cf(y) = \int_{\Gamma\cap N\backslash N} f(ny)dn.$$

Lemma 3.7.1 (Properties of Incomplete Theta Series). *Let θ_0 denote the closed subspace of $L^2(SL(2,\mathbb{Z})\backslash H)$ generated by $T\psi$ such that $M\psi(-1) = (T\psi, 1) = 0$, where $T\psi$ is the incomplete theta series (3.132) and ψ is smooth with compact support on $\mathbb{R}^+$and Mellin transform*

$$M\psi(s) = \int_0^\infty \psi(y)y^{s-1}dy.$$

Define

$$L_0^2(SL(2,\mathbb{Z})\backslash H) = \left\{ f \in L^2(SL(2,\mathbb{Z})\backslash H) \;\middle|\; (Cf)(y) = 0 \quad \text{for almost all } y > 0 \right\}.$$

Here $C(y)$ is the constant term operator in Exercise 3.7.3(a). Then we have the following list of assertions.

(1) There is an orthogonal decomposition

$$L^2(SL(2,\mathbb{Z})\backslash H) = L_0^2(SL(2,\mathbb{Z})\backslash H) \oplus \mathbb{C} \oplus \theta_0.$$

(2) For $c > 1$ and $E_s(z)$ the Eisenstein series defined by formula (3.81),

$$T\psi(z) = \frac{1}{2\pi i}\int_{\mathrm{Re}\, s=c} M\psi(-s)E_s(z)ds.$$

(3) If $\varphi(s) = \Lambda(1-s)/\Lambda(s)$ and $\Lambda(s) = \pi^{-s}\Gamma(s)\zeta(2s)$, then

$$(T\psi, E_s) = M\psi(\bar{s}-1) + \varphi(\bar{s})M\psi(-\bar{s}).$$

(4) Set $v_0^2 = (\mathrm{area}(SL(2,\mathbb{Z})\backslash H))^{-1} = \mathrm{Res}_{s=1}E_s(z)$. Then v_0 is in the discrete part of the spectral decomposition of Δ on $SL(2,\mathbb{Z})\backslash H$, and

$$T\psi(z) = (T\psi, v_0)v_0 + \frac{1}{4\pi i}\int_{\mathrm{Re}\, s=\frac{1}{2}} (T\psi, E_s)E_s(z)ds.$$

Proof.

(1) From Exercise 3.7.3, we have

$$(T\psi, g) = \int_{y>0} \psi(y)\overline{(Cg)(y)}y^{-2}dy,$$

where $C(g)$ is the constant term in the Fourier expansion of g. It follows that the orthogonal complement of the L^2-space spanned by the $T\psi$s is the space $L^2_0(SL(2,\mathbb{Z})\backslash H)$.

Next use Exercise 3.7.3 to see that

$$(T\psi, 1) = M\psi(-1).$$

This means that we can decompose the L^2-span of the incomplete theta series into an orthogonal direct sum of the constants and Θ_0. To see why the constants come out discretely in the spectral decomposition, see the argument in the proof of part (4).

(2) From the Mellin inversion formula of Sect. 1.4 it follows that

$$\psi(y) = \frac{1}{2\pi i}\int_{\mathrm{Re}\, s=c} M\psi(-s)y^s ds.$$

In order to prove (2), use the definitions of $T\psi$ in Eq. (3.132) and the Eisenstein series E_s in formula (3.81) and sum this Mellin inversion formula over $\Gamma_\infty\backslash\Gamma$ for $\mathrm{Re}\, s = c > 1$. This last condition is necessary for absolute convergence.

(3) This follows easily from Exercise 3.7.3 and the Fourier expansion of E_s given in Exercise 3.5.4 of Sect. 3.5.

(4) For $c > 1$, we have from part (2),

$$2T\psi(z) = \frac{1}{2\pi i}\int_{\mathrm{Re}\, s=c} 2M\psi(-s)E_s(z)ds.$$

To use the residue theorem and move the integral over to the line $\mathrm{Re}\, s = \frac{1}{2}$, we must find the poles of the integrand in $\frac{1}{2} \le \mathrm{Re}\, s \le c$. Now $M\psi(-s)$ is entire, since ψ has compact support. Thus the only pole of the integrand in the region is at $s = 1$ with residue

$$2M\psi(-1)\mathrm{Res}_{s=1}E_s(z) = 2(T\psi, 1)(\mathrm{vol}(\Gamma\backslash H))^{-1}.$$

So we have

$$2T\psi(z) = 2(T\psi, v_0)v_0 + \frac{1}{2\pi i}\int_{\mathrm{Re}\, s=\frac{1}{2}} 2M\psi(-s)E_s(z)ds.$$

Now we use the functional equation of E_s (noting that $|\varphi(s)| = 1$, for $\mathrm{Re}\, s = \frac{1}{2}$) to write

$$
\begin{aligned}
&\int_{\mathrm{Re}\,s=\frac{1}{2}} 2M\psi(-s)E_s(z)ds \\
&= \int_{\mathrm{Re}\,s=\frac{1}{2}} \left(M\psi(-s)E_s(z)+\varphi(s)M\psi(-s)E_{1-s}(z)\right) ds \\
&= \int_{\mathrm{Re}\,s=\frac{1}{2}} \left(M\psi(-s)E_s(z)+\varphi(1-s)M\psi(-(1-s))E_s(z)\right) ds.
\end{aligned}
$$

The last equality follows from replacing s by $1-s$ in the second part of the integrand.

Now note that when $\mathrm{Re}\,s=\frac{1}{2}$, we have $-s=\overline{s}-1$ and $1-s=\overline{s}$. Thus, it follows from part (3) that

$$
2\,T\psi(s) = 2(T\psi, v_0)v_0 + \frac{1}{2\pi i}\int_{\mathrm{Re}\,s=\frac{1}{2}} (T\psi, E_s)E_s(z)ds,
$$

as was to be proved. □

Exercise 3.7.4. Fill in the details in the proofs of parts (2)–(4) of Lemma 3.7.1 above. For part (4), note that you need bounds on the integrand to move the lines of integration. The argument is similar to that used for ordinary Mellin transforms.

Exercise 3.7.5 (The Size of $E_s(z)$ as t Approaches Infinity and the Residue at $s=1$).

(a) Show that $E_s(z)=O(t^{\frac{1}{2}+\varepsilon})$, as t approaches infinity (for fixed z), if $s=\sigma+it$ and $\sigma\geq\frac{1}{2}$.

Hint. *Use the Phragmen–Lindelof theorem (see Lang [388, pp. 262–267] or Titchmarsh [680]). Note that* $E_s(z)=O(\exp(t^a))$ *for some* $a>0$*, using the incomplete gamma expansion (Theorem 1.4.1 of Sect. 1.4) or the Fourier series (Exercise 3.5.4 of Sect. 3.5). Deduce that for* $\mathrm{Re}\ s>1$, $E_s(z)=O(1)$*, using the definition of* $E_s(z)$ *as a Dirichlet series. Obtain the estimate for* $E_s(z)$ *when* $\mathrm{Re}\ s>0$*, using the functional equation*

$$
E_s(z)=\frac{\Lambda(1-s)}{\Lambda(s)}E_{1-s}(z)=\frac{\Lambda(1-s)}{\Lambda(\frac{1}{2}-s)}E_{1-s}(z), \quad \text{with } \Lambda(s)=\pi^{-s}\Gamma(s)\zeta(2s),
$$

and using Stirling's formula for $\Gamma(s)$*. One could actually improve the estimate, making use of Hadamard's three lines theorem, and one could do even better if one knew the Riemann hypothesis.*

(b) Show that $\mathrm{Res}_{s=1}E_s(z)=(\mathrm{area}(SL(2,\mathbb{Z})\backslash H))^{-1}$. You could use the Fourier expansion of E_s to do this, for example, plus the fact that $\zeta(2)=\pi^2/6$. Or you could use the incomplete gamma expansion of Epstein's zeta function (Theorem 1.4.1 of Sect. 1.4). This is a general property of Eisenstein series as we shall see in Volume II or [667] (see also Sarnak [558] and Siegel [600, Vol. I, pp. 459–468; Vol. III, pp. 328–333]).

Lemma 3.7.1 completes the portion of the proof of Theorem 3.7.1 coming from the continuous spectrum and the one-dimensional space of constants in the discrete spectrum. This latter part of the discrete spectrum is often called the "residual spectrum" since it comes from a residue of the Eisenstein series. This use of the words "residual spectrum" conflicts with that of functional analysis. There are groups Γ such that the Eisenstein series have arbitrarily many poles on the real line (see Selberg [571, p. 184]). Next we must deal with the rest of the discrete spectrum of Δ. We will show that $L_0^2(SL(2,\mathbb{Z})\backslash H)$ lies in the discrete spectrum of Δ (see Theorem 3.7.2 below).

At this point one might wonder why $L_0^2(SL(2,\mathbb{Z})\backslash H)$ is not the zero space. This is equivalent to showing that nonzero cusp forms exist. One can deduce this from the Selberg trace formula itself. It is fairly easy to show that there are odd cusp forms for $SL(2,\mathbb{Z})$, as the following exercise shows. See our discussion of the Roelcke–Selberg conjecture in Sect. 3.5.5.

Exercise 3.7.6.

(a) Show that $L_0^2(SL(2,\mathbb{Z})\backslash H)$ contains a nonzero element.
(b) What goes wrong with trying to find an even element of $L_0^2(SL(2,\mathbb{Z})\backslash H)$ by starting with a classical holomorphic cusp form of weight k and forming $g(z) = y^{k/2}|f(z)|$?

Hint on Part (a) (H. Stark). *Take an odd function $g(x)$ of period 1; e.g., $\sin(2\pi nx)$, $x \in \mathbb{R}$, $n \in \mathbb{Z}$. Then form a function $f(z)$ on the standard fundamental domain D for $SL(2,\mathbb{Z})$ pictured in Fig. 3.14 by setting*

$$f(z) = \begin{cases} g(x) & \text{for } z = x+iy, \text{ with } \; y \geq 2, \\ 0 & \text{for } z = x+iy, \text{ with } \; y < 2. \end{cases}$$

We could actually do this smoothly. Then extend $f(z)$ to a function $F(z)$ by writing $F(z) = f(\gamma z)$, if $\gamma \in SL(2,\mathbb{Z})$ is such that $\gamma z \in D$. Note that the lack of uniqueness of γ for z on the boundary of D does not matter since f is the same at equivalent boundary points. You need to show that the function $F(z)$ is an odd function of x. This comes from the fact that $f(z)$ was an odd function of x and the identity

$$\gamma^*(-\bar{z}) = -\overline{\gamma z} \quad \text{for } \gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL(2,\mathbb{Z}) \quad \text{and } \gamma^* = \begin{pmatrix} a & -b \\ -c & d \end{pmatrix}.$$

In order to study the discrete spectrum of Δ it is natural to attempt to replace Δ with integral operators. Here we shall use convolution operators. It is also possible to study the **resolvent**:

$$(\Delta - \lambda)^{-1} f(z) = \int\limits_{\substack{w \in \Gamma\backslash H \\ w=u+iv}} G_\lambda(z,w) f(w) \frac{du\, dv}{v^2}. \tag{3.133}$$

The kernel $G_\lambda(z,w)$ in formula (3.133) is called the **Green's function** for $\Delta - \lambda$. We studied the Euclidean analogue of this Green's function in formula (1.16). References for the non-Euclidean Green's function are Elstrodt [155, 156], Faddeev [171], Fay [176], Hejhal [261], Lang [389], Neuenhöffer [494], and Roelcke [544, 545]. The poles in λ of the Green's function G_λ are the eigenvalues of Δ on $L^2(SL(2,\mathbb{Z})\backslash H)$; i.e., the discrete spectrum. Branching in λ for G_λ gives rise to the continuous spectrum and the Kodaira–Titchmarsh formula allows the computation of the spectral measure from the jump across the real axis [see formula (2.26)]. See Example 3.7.2 and Exercises 3.7.10–3.7.15 of this section for more information on Green's functions.

In order to define convolution as in Sect. 3.2 we must think of functions on $H \cong G/K$, with $G = SL(2,\mathbb{R})$ and $K = SO(2)$, as functions on G by writing $f(a) = f(ai)$. Here ai means the point in H to which $a \in SL(2,\mathbb{R})$ maps $i = \sqrt{-1}$ by fractional linear transformation (see Exercise 3.1.8 of Sect. 3.1). Suppose then that we have two integrable functions $f, g : H \to \mathbb{C}$. Then define the **convolution** of f and g by

$$L_g f(a) = (f * g)(a) = \int_G f(b) g(b^{-1}a)\, db. \tag{3.134}$$

Here db denotes left = right Haar measure on G (see Sect. 2.1 and Exercise 3.2.19 of Sect. 3.2). The measure db is unique up to a positive constant and is characterized by its invariance under the group action. General results on Haar measure can be found in Helgason [277] and Lang [387]. We do not need any explicit formula for db, just its existence.

Note. There is a slightly confusing aspect of the identification $f(ai) = f(a)$, for $a \in SL(2,\mathbb{R})$. For if we adjust the Haar measure da so that

$$\int_G f(ai)\,da = \int_H f(x+iy)\, y^{-2} dx\, dy,$$

we discover quickly that the integral on the left is invariant under the substitution of a^{-1} for a, but not the integral on the right! For

$$x + iy = \begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix} \begin{pmatrix} y^{1/2} & 0 \\ 0 & y^{-1/2} \end{pmatrix};$$

and

$$\left(\begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix} \begin{pmatrix} y^{1/2} & 0 \\ 0 & y^{-1/2} \end{pmatrix} \right)^{-1} i = \begin{pmatrix} y^{-1/2} & 0 \\ 0 & y^{1/2} \end{pmatrix} \begin{pmatrix} 1 & -x \\ 0 & 1 \end{pmatrix} i = \frac{i - x}{y}.$$

Let $u + iv = (i - x)/y$. The Jacobian of this change of variables is

$$\frac{\partial(u,v)}{\partial(x,y)} = \frac{1}{y^3}.$$

Therefore

$$\int_H f\left(\left(\begin{pmatrix}1 & x\\ 0 & 1\end{pmatrix}\begin{pmatrix}y^{1/2} & 0\\ 0 & y^{-1/2}\end{pmatrix}\right)^{-1}\right) y^{-2}dx\,dy = \int_H f(u+iv)v^{-1}du\,dv.$$

This is a reflection of the non-unimodularity of the group of upper triangular matrices with determinant 1 and positive diagonal. The moral is that we should be somewhat careful about our identifications.

As an example, let us compute $L_g f(i)$, for $f(z) = p_s(z) = (\operatorname{Im} z)^s$. The result is

$$\begin{aligned} L_g p_s(i) &= \int_G p_s(bi)g(b^{-1})db = \int_H y^s g\left(\frac{i-x}{y}\right) y^{-2}dx\,dy \\ &= \int_H y^{-s}g(z)y^{-1}dx\,dy = \hat{g}(1-\overline{s}), \end{aligned}$$

where $\hat{g}(s)$ is the Helgason transform of g defined in formula (3.32) of Sect. 3.2, assuming that g is in fact rotation-invariant.

Lemma 3.7.2 (Properties of Convolution Operators). *Suppose throughout that* $g : H \to \mathbb{C}$ *is infinitely differentiable with compact support.*

(1) *The operator* L_g *defined by* Eq. (3.134) *commutes with the left action of* $x \in G$ *on functions. Thus* L_g *is a* G*-invariant integral operator such that*

$$L_g : L^2(SL(2,\mathbb{Z})\backslash H) \to C^\infty(SL(2,\mathbb{Z})\backslash H).$$

(2) *If* $g(a) = g(a^{-1})$ *for all a in* G *and if* g *is real-valued, then* L_g *is a self-adjoint operator on* $L^2(SL(2,\mathbb{Z})\backslash H)$; i.e., $(L_g f, h) = (f, L_g h)$, *using the inner product* (3.131).

(3) $L_{g*h} f = L_g L_h f$.

(4) *The operators* L_g *commute if* g *is* K*-bi-invariant*, $K = SO(2)$; that is *if* $g : K\backslash G/K \to \mathbb{C}$.

(5) *If* $g(a) = g(a^{-1})$ *for all* $a \in G$, *then* $\Delta(L_g) = L_g\Delta$ *and* $\Delta L_g = L_{\Delta g}$.

Proof.

(1) For x in G, set $f^x(y) = f(xy)$. Then, using the left invariance of Haar measure db, we obtain

$$\begin{aligned} L_g(f^x)(a) &= \int_G f(xb)g(b^{-1}a)\,db \\ &= \int_G f(b)g(b^{-1}xa)\,db = (L_g f)^x(a). \end{aligned}$$

(2) To see this, write

$$\begin{aligned}
(L_g f, h) &= \int_{\Gamma\backslash G/K} \int_{G/K} f(b) g(b^{-1}a)\, db\, \overline{h(a)}\, da \\
&= \int_{\Gamma\backslash G/K} \int_{\Gamma\backslash G/K} f(b) \sum_{\gamma\in\Gamma} g(b^{-1}\gamma a)\, db\, \overline{h(a)}\, da \\
&= \int_{G/K} \int_{G/K} g(b^{-1}a) \overline{h(a)}\, da\, f(b)\, db \\
&= (f, L_g h), \text{ since } g(a) = g(a^{-1}) \text{ and } g \text{ is real-valued.}
\end{aligned}$$

(3) This follows from the associative law for the convolution of L^1 functions.
(4) We leave this as an Exercise similar to Exercise 3.2.19.
(5) See Exercise 3.7.7.

□

Exercise 3.7.7. Prove part (5) of Lemma 3.7.2. Do both formulas require $g(a) = g(a^{-1})$? If g is K-bi-invariant, does it follow that $g(a) = g(a^{-1})$?

Lemma 3.7.3 (The Connection between the Eigenvalues of Convolution Operators and Eigenvalues of the Laplacian. The Selberg Transform and the Helgason Transform).

(a) *Suppose that $L_g f = \lambda_g f$, $\lambda_g \in \mathbb{C}$, for all $g \in C_c^\infty(K\backslash G/K)$. Then $\Delta f = \mu f$, and the transform $\mu \to \lambda_g$, which is often called the Selberg transform, is essentially the Helgason transform defined by formula (3.32) for rotation-invariant functions g.*

(b) *More precisely, suppose that $g : K\backslash G/K \to \mathbb{C}$ is infinitely differentiable with compact support. Let ϕ be any eigenfunction of the non-Euclidean Laplacian with $\Delta\phi = s(s-1)\phi$. For example, ϕ could be a cusp form, or y^s, or the spherical function*

$$h_s(z) = \int_{k\in K} (\mathrm{Im}(kz))^s dk = P_{s-1}(\cosh r), \quad \textit{if } z = ke^{-r}i \in H, k \in K,$$

with P_{s-1} = the Legendre function of Exercises 3.2.9 and 3.2.10 of Sect. 3.2. Then $\Delta\phi = s(s-1)\phi$ implies that

$$\begin{aligned}
\frac{L_g\phi}{\phi}(x) = \frac{\phi * g}{\phi}(x) &= \hat{g}(1-\bar{s}) = \int_H g(z) y^{-s-1} dx\, dy \\
&= \text{the Helgason transform of } g;
\end{aligned}$$

i.e., if the eigenvalue of Δ corresponding to ϕ is $\mu = s(s-1)$, then the eigenvalue of L_g corresponding to ϕ is $\lambda_g = \hat{g}(1-\bar{s})$.

Proof.

(a) Suppose that $f * g = \lambda_g f$. Then

$$\lambda_g \Delta f = \Delta(f * g) = f * (\Delta g) = \lambda_{\Delta g} f \quad \text{implies } \Delta f = \mu f,$$

with $\mu = \lambda_{\Delta g}/\lambda_g$. Note that if g runs through a Dirac family at the identity, we can assume that λ_g is nonzero.

Question. Does it suffice to know that $L_g f = \lambda_g f$ for all g in a Dirac family at the identity?

(b) Let $\mathcal{A}$ be the operator that averages over the compact group $K = SO(2)$. Then

$$\mathcal{A}\phi(x) = \int_K \phi(kx)dk,$$

where dk is left (which equals right) Haar measure on $K = SO(2)$, normalized so that

$$\int_K dk = 1.$$

We have proved the uniqueness of the spherical function $h_s(z) = P_{s-1}(\cosh r)$, for $z = ke^{-r}i$, with $k \in K$, $r > 0$. See formula (3.25) of Sect. 3.2 with $a = 0$ and Exercise 3.2.9 of that section. Thus $\mathcal{A}\phi = ch_s$ with $c = \phi(i)$, for $i = \sqrt{-1}$. It follows that

$$\mathcal{A}(\phi * g)(a) = ((\mathcal{A}\phi) * g)(a) = ((ch_s) * g)(a) = c(h_s * g)(a).$$

Evaluate this at $a =$ the identity, I, to obtain:

$$L_g\phi(I) = (\phi * g)(I) = c(h_s * g)(I) \quad \text{where } c = \phi(I) = \phi(i),$$

by our identification. Therefore we can set $f(z) = y^s$ and obtain

$$\lambda_g = \frac{L_g f}{f}(i) = \frac{L_g \phi}{\phi}(i) = \int_H g(z) y^{-s-1} dxdy = \hat{g}(1-\bar{s}),$$

as we saw in the note after the definition of the convolution operator in formula (3.134).

Note that if $\Delta f = \mu f$, then $L_g f = \lambda_g f$. For if $a \in G$, set $f^a(x) = f(ax)$, when $x \in G$. Then $\Delta(f^a) = (\Delta f)^a = \mu f^a$. So, using the preceding results, we find that:

$$L_g(f^a)(I) = (L_g f)^a(I) = \hat{g}(1-\bar{s})(f^a)(I),$$

which implies that $(L_g f)(a) = \hat{g}(1-\bar{s}) f(a)$. □

We are now in a position to prove the remaining part of Theorem 3.7.1.

Theorem 3.7.2. *Suppose $g : H \to \mathbb{C}$ is infinitely differentiable with compact support. Let L_g be the convolution operator in Eq. (3.134). Then L_g is a compact operator on the space $L_0^2(\Gamma\backslash H)$, $\Gamma = SL(2,\mathbb{Z})$, defined in Lemma 3.7.1. Thus the spectrum of Δ on $L_0^2(\Gamma\backslash H)$ is discrete and the spectral theorem for compact, self-adjoint operators (see Theorem 1.3.7 of Sect. 1.3) assures the existence of a complete orthonormal set of cusp forms spanning $L_0^2(\Gamma\backslash H)$.*

Proof. Let $g : K\backslash G/K \to \mathbb{C}$ be C^∞ with compact support for $G = SL(2,\mathbb{R})$, $K = SO(2)$. We want to show that L_g is a compact operator on $L_0^2(\Gamma\backslash H)$. Suppose that $\|\,\|$ denotes the sup norm for functions on G; i.e.,

$$\|f\| = \sup\{|f(x)|,\ x \in G\}.$$

And let $\|\,\|_2$ denote the L^2 norm for functions on $\Gamma\backslash H$; i.e.,

$$\|f\|_2 = (f,f)^{1/2} \text{ with the inner product in } Eq.\,(3.131).$$

We want to show that there is a positive constant C such that for any f in $L_0^2(\Gamma\backslash H)$,

$$\|L_g f\| \le C\|f\|_2. \tag{3.135}$$

This will imply that L_g is a compact operator on $L_0^2(\Gamma\backslash H)$. To see this, note that the image of L_0^2 under L_g is an equicontinuous family of continuous functions. If we can show the family to be uniformly bounded, then the theorem of Arzelà-Ascoli (see Kolmogorov and Fomin [366]) implies that sequences $\{L_g f_n\}$ must have subsequences converging on compacta in $\Gamma\backslash H$ to continuous functions. Lang [389] gives another proof of this.

Now, in order to bound the $L_g f$s by the L_2 norms of the $f's$, we argue as follows, using the Poisson summation formula from Sect. 1.3. Assuming that $g(a) = g(a^{-1})$, we can write

$$L_g f(a) = \int_H f(z) g(a^{-1}z) y^{-2} dx\, dy.$$

Suppose next that

$$a^{-1} = k\begin{pmatrix} d^{1/2} & 0 \\ 0 & d^{-1/2}\end{pmatrix}\begin{pmatrix} 1 & v \\ 0 & 1\end{pmatrix} \quad \text{with } k \in K,\ d > 0,\ v \in \mathbb{R}.$$

It is possible to make this decomposition by Exercise 3.1.9 of Sect. 3.1. In the general case considered in Volume II [667], this is called the **Iwasawa decomposition** of a in G.

We are trying to show that $L_g f(a)$ is bounded as $ai = -v + i/d$ approaches infinity in H; i.e., as v approaches $+\infty$ or $-\infty$, or d approaches 0 or infinity. We will assume that $g(x+iy) = h(x)k(y)$ to simplify the calculations, since g can be approximated by such functions. Now set

$$\Gamma_\infty = \left\{ \begin{pmatrix} 1 & n \\ 0 & 1 \end{pmatrix} \middle| \, n \in \mathbb{Z} \right\}.$$

Then

$$\begin{aligned} L_g f(a) &= \int_{\Gamma_\infty \backslash H} \sum_{n \in \mathbb{Z}} g(d(z+n+v)) f(z) y^{-2} \, dx \, dy \\ &= \int_{\Gamma_\infty \backslash H} \sum_{n \in \mathbb{Z}} h(d(x+n+v)) k(dy) f(z) y^{-2} \, dx \, dy. \end{aligned}$$

Poisson tells us (using the dilation and translation properties of the Fourier transform on $\mathbb{R}$) that

$$\sum_{n \in \mathbb{Z}} h(d(x+n+v)) = d^{-1} \sum_{n \in \mathbb{Z}} \hat{h}(n/d) \exp[2\pi i n(x+v)],$$

where $\hat{h}$ denotes the Fourier transform of h over $\mathbb{R}$. The fact that h is C^∞ with compact support implies that $|\hat{h}(x)| \leq c|x|^{-p}$ for any power p. So if $n \neq 0$, we have a bound. If $n = 0$, since f is in L_0^2, we find that

$$\int_{\Gamma_\infty \backslash H} \hat{h}(0) k(dy) f(z) y^{-2} \, dx \, dy = 0.$$

Thus we have

$$|L_g f(a)| \leq c d^{p-1} \int_{\Gamma_\infty \backslash H} \sum_{n \neq 0} |n|^{-p} |k(dy)| |f(z)| y^{-2} \, dx \, dy.$$

We have assumed that the function k has compact support. Thus for d small enough, the preceding integral will be bounded by a constant times

$$d^{p-1} \int_{\substack{|x| \leq 1/2 \\ y > 2}} |f(z)| y^{-2} \, dx \, dy.$$

The Cauchy–Schwarz inequality completes the proof of the desired inequality (3.135). Finally, the proofs of Theorems 3.7.1 and 3.7.2 are complete. □

Question. Suppose we replace $SL(2,\mathbb{Z})$ in Theorem 3.7.1 by some congruence subgroup $\Gamma(N)$. Can you formulate harmonic analysis on $\Gamma(N) \backslash H$ as some sort of product of that on $SL(2,\mathbb{Z}) \backslash H$ with that on the finite group $SL(2,\mathbb{Z})/\Gamma(N)$?

3.7.3 A Non-Euclidean Poisson Summation Formula

It is now possible to develop a non-Euclidean analogue of the Poisson summation formula, using the same argument that proved Theorem 1.3.2 of Sect. 1.3.

Theorem 3.7.3 (A Non-Euclidean Poisson Summation Formula). *Let* $f : H \to \mathbb{C}$ *be* $K = SO(2)$*-invariant; i.e.,* $f(kz) = f(z)$ *for all* $k \in K$, $z \in H$ *and suppose also that for* $\Gamma = SL(2,\mathbb{Z})$*, the series*

$$g(z) = \sum_{\gamma \in \Gamma/\{\pm I\}} f(\gamma z)$$

converges absolutely and uniformly on compacta in H *to a function* $g(z)$ *such that* $\Delta g \in L^2(\Gamma \backslash H)$*. Then, using the notation (3.129) and (3.130),*

$$g(z) = \sum_{n \geq 0} \hat{f}(s_n)\overline{v_n(i)}v_n(z) + \frac{1}{4\pi i}\int_{\mathrm{Re}\,s=\frac{1}{2}} \hat{f}(s)\overline{E_s(i)}E_s(z)ds,$$

with $\hat{f}(s)$ *=the non-Euclidean Fourier (or Helgason transform) given by*

$$\hat{f}(s) = \int_H f(z)\overline{y^s}\, y^{-2}\, dx\, dy.$$

Proof. By Theorem 3.7.1, we can write

$$g(z) = \sum_{n \geq 0} (g, v_n)v_n(z) + \frac{1}{4\pi i}\int_{\mathrm{Re}\,s=\frac{1}{2}} (g, E_s)E_s(z)ds.$$

Convergence is uniform and absolute by Exercise 3.7.17 that follows. Using the fact that H is a disjoint union of translates by $\gamma \in \Gamma/\{\pm I\}$ of the fundamental domain $\Gamma \backslash H$, we see that

$$(g, v_n) = \int_{\Gamma \backslash H} \sum_{\gamma \in \Gamma/\{\pm I\}} f(\gamma z)\overline{v_n(z)}\, y^{-2}\, dx\, dy = \int_H f(z)\overline{v_n(z)}\, y^{-2}\, dx\, dy.$$

Then Lemma 3.7.3 implies that $(g, v_n) = \hat{f}(s_n)\overline{v_n(i)}$. The formula with v_n replaced by E_s is proved similarly, to complete the proof of Theorem 3.7.3. □

Later, while discussing Selberg's trace formula, we shall find another derivation of Theorem 3.7.3 (from the non-Euclidean analogue of Mercer's theorem which was Theorem 1.3.8 of Sect. 1.3).

The analogue of Theorem 3.7.3, for Γ such that $\Gamma \backslash H$ is compact, goes back to Delsarte in 1942. Delsarte used this result to study non-Euclidean lattice point problems such as the asymptotic result given in formula (3.136) below. Others have used Theorem 3.7.3 to study Green's functions for the Laplacian on

$\Gamma\backslash H$ (see Exercise 3.7.14 below). It is also possible to use the non-Euclidean Poisson summation formula to study the Poincaré series that arise in work on the Ramanujan–Petersson conjecture for Fourier coefficients of modular forms (see Bruggeman [68], Deshouillers and Iwaniec [130], Goldfeld and Sarnak [221], Kuznetsov [381], and Selberg [570]). Kudla and Millson [379] use non-Euclidean Poisson summation to give an explicit construction of the harmonic one-form dual to an oriented closed geodesic on an oriented Riemann surface of genus greater than 1. Takhtadzhyan and Vinogradov [654] use such results to obtain relations between fundamental units of real quadratic fields and the existence of cusp forms with eigenvalue $\frac{1}{4}$ for $\Gamma_0(d)$.

Example 3.7.1 (Non-Euclidean Lattice Point Problems). References include Buser [74], Delsarte [129, Tome II, pp. 599–601, 829–845], Huber [306], Mennicke [463], Nicholls [495], and Patterson [510]. Recall first that in Theorem 1.3.5 of Sect. 1.3 we found an asymptotic result for the number of lattice points; i.e., elements of $\mathbb{Z}^2$, in a circle, as the radius of the circle blows up. One non-Euclidean analogue of this is

$$\left.\begin{array}{c} N_\Gamma(x) = \#\{\gamma \in \Gamma/\{\pm I\} \mid \cosh d(i,\gamma i) \leq x\} \\ \sim \frac{2\pi x}{\text{area}(\Gamma\backslash H)}, \text{ as } x \to \infty, \end{array}\right\} \tag{3.136}$$

for $\Gamma = SL(2,\mathbb{Z})$, where $d(z,w)$ denotes the non-Euclidean distance between points $z,w \in H$. Note that $2\pi/\text{area}(\Gamma\backslash H) = 6$ in this case.

You might be worrying about the hyperbolic cosine in the definition of $N_\Gamma(x)$ in formula (3.136). Hopefully the following exercise will allay these worries by indicating that there is a generalization to $SL(n,\mathbb{Z})$ or $GL(n,\mathbb{Z})$ (which will be considered in Volume II or [667]).

Exercise 3.7.8.

(a) Recall Exercises 3.1.9 and 3.1.10 of Sect. 3.1. These exercises show that if $z = ke^{-r}i$, for $k \in K = SO(2)$, $r > 0$, and W_z is the corresponding 2×2 matrix in $\mathcal{SP}_2$, then

$$W_z = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \left[\begin{pmatrix} e^{-r/2} & 0 \\ 0 & e^{r/2} \end{pmatrix} k \right],$$

Here we use the notation $Y[A] = {}^tAYA$, with ${}^tA =$ transpose of A. Show that

$$\cosh d(i,z) = \frac{1}{2}(e^r + e^{-r}) = \frac{1}{2}\text{Tr}(W_z).$$

Then show that if

$$\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma,$$

$$2\cosh d(\gamma i, i) = \mathrm{Tr}\ ({}^t\gamma\gamma) = a^2+b^2+c^2+d^2.$$

(b) Prove that $2\cosh d(z,w)+2 = \frac{|z-\overline{w}|^2}{\mathrm{Im}\, z\, \mathrm{Im}\, w}$.

Hint. *See Exercise 3.2.25 of Sect. 3.2.*

From Exercise 3.7.8, we see that formula (3.136) is equivalent to the following asymptotic formula:

$$\#\left\{\gamma \in \Gamma/\{\pm I\} \;\middle|\; \frac{1}{2}\mathrm{Tr}({}^t\gamma\gamma) \le x\right\} \sim 6x, \quad \text{as} \quad x \to \infty,$$

which means that

$$\#\left\{a,b,c,d \in \mathbb{Z} \;\middle|\; a^2+b^2+c^2+d^2 \le x,\ ad-bc=1\right\} \sim 6x, \ \text{ as } \ x \to \infty. \tag{3.137}$$

We know from Theorem 1.3.5 of Sect. 1.3 that

$$\#\left\{a,b,c,d \in \mathbb{Z} \;\middle|\; a^2+b^2+c^2+d^2 \le x\right\} \sim \pi^2x^2, \ \text{ as } \ x \to \infty. \tag{3.138}$$

A comparison of these two asymptotic results gives one a good feeling for the relative densities of these two sets.

In order to prove Eq. (3.136), one may imitate the proof of Theorem 1.3.5 of Sect. 1.3. To this end, we form the **non-Euclidean theta function** [12] for $a > 0$:

$$\theta_\Gamma(a) = \sum_{\gamma\in\Gamma/\{\pm I\}} \exp\left[-\frac{1}{2}a\mathrm{Tr}({}^t\gamma\gamma)\right] = \sum_{\gamma\in\Gamma/\{\pm I\}} \exp[-a\cosh d(i,\gamma i)]. \tag{3.139}$$

Set

$$f_a(z) = \exp[-a\cosh d(i,z)] = \exp\left[-\frac{1}{2}a\ \mathrm{Tr}(W_z)\right] \quad \text{for } z \in H,$$

where W_z is the corresponding matrix in $\mathcal{SP}_2$ from Exercise 3.1.9 given for $z = x+iy$ by

$$W_z = \begin{pmatrix} 1/y & 0 \\ 0 & y \end{pmatrix}\begin{bmatrix} 1 & x \\ 0 & 1 \end{bmatrix} = \begin{pmatrix} 1/y & * \\ * & (x^2+y^2)/y \end{pmatrix}. \tag{3.140}$$

The Helgason transform is

$$\widehat{f_a}(s) = \int\limits_{W_z\in\mathcal{SP}_2} \exp\left[-\frac{1}{2}a\ \mathrm{Tr}\ (W_z)\right]\ y^{\bar{s}-2}dx\,dy = 2\sqrt{\frac{2\pi}{a}}K_{\bar{s}-\frac{1}{2}}(a). \tag{3.141}$$

---

[12] Our non-Euclidean theta function in Eq. (3.139) is not the same as that considered by differential geometers (see Molchanov [473, p. 46] and the proof of Theorem 3.7.5).

To prove Eq. (3.141), use formula (3.140) to see that $\mathrm{Tr}(W_z) = \left(1+(x^2+y^2)\right)/y$, and

$$\widehat{f}_a(s) = \int\limits_{y>0}\int\limits_{x\in\mathbb{R}} \exp\left[-a\left(\frac{1+x^2+y^2}{2y}\right)\right] y^{\bar{s}-2}\,dx\,dy.$$

Perform the integral over x and obtain

$$\int\limits_{x\in\mathbb{R}} \exp\left(\frac{-ax^2}{2y}\right)\,dx = \sqrt{\frac{2\pi y}{a}}.$$

It follows that

$$\widehat{f}_a(s) = \sqrt{\frac{2\pi}{a}} \int\limits_{y>0} y^{\bar{s}-\frac{1}{2}} \exp\left[-\frac{a}{2}\left(y+\frac{1}{y}\right)\right]\,\frac{dy}{y}.$$

A formula from Exercise 3.2.1 of Sect. 3.2 finishes the proof of Eq. (3.141).

Applying Theorem 3.7.3 to $f_a(z) = \exp[-\frac{1}{2}a\,\mathrm{Tr}(W_z)]$, we have

$$\begin{aligned}\theta_\Gamma(a) &= \sum_{\gamma\in\Gamma/\{\pm I\}} \exp\left[-\frac{a}{2}\mathrm{Tr}({}^t\gamma\gamma)\right]\\ &= 2\sqrt{\frac{2\pi}{a}}\sum_{n\geq 0} K_{\bar{s}_n-\frac{1}{2}}(a)|v_n(i)|^2 + 2\sqrt{\frac{2\pi}{a}}\frac{1}{4\pi i}\int_{\mathrm{Re}\,s=\frac{1}{2}} K_{\bar{s}-\frac{1}{2}}(a)\,|E_s(i)|^2 ds.\end{aligned} \tag{3.142}$$

In order to obtain the asymptotics of $\theta_\Gamma(a)$, as $a\to 0+$, we need the following results:

$$K_{iu}(a) = \sqrt{\frac{2\pi}{u}}e^{-u\pi/2}\left(\cos\left[u\log\frac{a}{2} - u\log u + u + \frac{\pi}{4}\right] + o(1)\right),\ \text{as}\ \ u\to\infty,\ a\to 0; \tag{3.143}$$

$$E_{\frac{1}{2}+iu}(z) = O\left(|u|^{\frac{1}{2}+\varepsilon}\right) \quad \text{as}\ u\to\infty. \tag{3.144}$$

Formula (3.144) was Exercise 3.7.5 and formula (3.143) comes from Exercise 3.7.9 below.

One finds that the main term on the right-hand side of Eq. (3.142) is the term corresponding to v_0, and the rest is $O(1)$; i.e., bounded. It follows that

$$\theta_\Gamma(a) \sim \frac{2\pi}{\mathrm{area}(\Gamma\backslash H)a} \quad \text{as} \quad a\to 0+. \tag{3.145}$$

One finishes the proof of Eq. (3.136) with the same Tauberian theorem that sufficed to prove Theorem 1.3.5 of Sect. 1.3 (namely, Theorem 1.2.5 of Sect. 1.2).

Exercise 3.7.9. Prove formula (3.143) using the following facts about I and K Bessel functions (see Lebedev [401, p. 140]):

$$K_s(z) = \frac{\pi}{2}\frac{I_{-s}(z) - I_s(z)}{\sin(s\pi)} \quad \text{(taking limits if } s \in \mathbb{Z}\text{);}$$

$$I_s(z) = \left(\frac{z}{2}\right)^s \sum_{k=0}^{\infty} \frac{(z^2/4)^k}{k!\Gamma(s+k+1)}.$$

Example 3.7.2 (Green's Functions for $\Delta - \lambda$ on $\Gamma\backslash H$ by the Method of Images). References for this subject are Elstrodt [155,156], Faddeev [171], Fay [176], Hejhal [261,264, Vol. 2], Lang [389, Chap. 14], and Neuenhöffer [494]. We discussed the method of images briefly in Sect. 1.3. Here we seek a non-Euclidean analogue of the Green's function in Exercise 1.3.23 of Sect. 1.3.

The **Green's function** $G_\lambda(z,w)$ **for the resolvent** $R_\lambda = (\Delta - \lambda I)^{-1}$ **on the fundamental domain** $\Gamma\backslash H$, $\Gamma = SL(2,\mathbb{Z})$, satisfies

$$R_\lambda f(z) = \int_{w=u+iv\in\Gamma\backslash H} G_\lambda(z,w) f(w)\ v^{-2}\,du\,dv. \tag{3.146}$$

The **method of images** says that if $g_\lambda(z,w)$ is the kernel for the resolvent R_λ on all of H, then

$$G_\lambda(z,w) = \sum_{\gamma\in\Gamma/\{\pm I\}} g_\lambda(z,\gamma w). \tag{3.147}$$

We will show that $g_\lambda(z,w) = \frac{-1}{\pi} Q_{s-1}(\cosh d(z,w))$, if $\lambda = s(s-1)$ and Q_{s-1} denotes the Legendre function of the second kind. See formula (3.149).

So let us find $g_\lambda(z,w)$. In order to do this, one should recall the basic facts about Green's functions (see Courant and Hilbert [111, pp. 363–388], Garabedian [197] or Stakgold [618,619]). In particular, $g_\lambda(z,w)$ should have the same sort of singularity as that of the Euclidean Green's function for the Euclidean Laplacian on $\mathbb{R}^2$. The reason for this is that the y^2 in the non-Euclidean Laplacian is cancelled out by the y^{-2} in the G-invariant area element on H. This always happens when calculating Green's functions for differential operators with weights.

So the **singularity** of $g_\lambda(z,w)$ is the same as that of $(1/2\pi)\log|z-w|$. Also, $g_\lambda(z,w)$ must be $G = SL(2,\mathbb{R})$-invariant; i.e., $g_\lambda(az,aw) = g_\lambda(z,w)$ for all z,w in H and a in G. Moreover $g_\lambda(z,w)$ should be as small as possible as $z \to \infty$. Of course, g_λ should also satisfy the **differential equation**

$$y^2\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right) g_\lambda(z,w) = \lambda g_\lambda(z,w) \qquad \text{for } z \neq w.$$

Exercise 3.7.10.

(a) Show that the Green's function for $L = \partial^2/\partial x^2 + \partial^2/\partial y^2$ in $\mathbb{R}^2$ is $k(z,w) = (1/2\pi)\log|z-w|$; i.e.,

$$L^{-1}f(z) = \int_{w=u+iv\in\mathbb{R}^2} k(z,w)f(w)\,du\,dv.$$

Hint. *Use Green's theorem or Fourier transforms (see Courant and Hilbert [111], Garabedian [197], or Vladimirov [706]). Compare with Exercise 1.1.6 of Sect. 1.1 and Example 1.2.2 of Sect. 1.2.*

(b) Explain why the non-Euclidean Green's function $g_\lambda(z,w)$ for $(\Delta - \lambda)$ on H should have the same singularity as $k(z,w)$ as z approaches w.

Hint. *See Lang [389, pp. 276–280].*

Because $g_\lambda(z,w)$ is G-invariant, we can move w to i. Also, $g_\lambda(z,i)$ is a function of $d(z,i) = r =$ the geodesic radial coordinate of $z = ke^{-r}i$, for $k \in K = SO(2)$ and $r > 0$. Then, by separation of variables applied to

$$y^2\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right)g_\lambda(z,w) = \lambda g_\lambda(z,w) \quad \text{for} \quad z = x+iy \neq w = u+iv,$$

we find as in Exercise 3.2.11 of Sect. 3.2, that solutions which are functions of the radial variable alone have to be of the form

$$f(r) = AP_{-s}(\cosh r) + BQ_{-s}(\cosh r),\ \lambda = s(s-1).$$

Here $P_{-s}(u)$ and $Q_{-s}(u)$ are the **Legendre functions of the first and second kinds**, respectively (see Lebedev [401, Chap. 7]. In order to make $f(r)$ behave as $r \to \infty$, we choose $A = 0$. In order to obtain a singularity $(1/2\pi)\log|z-i|$ as $z \to i$ or $r \to 0$, we need the following exercise.

Exercise 3.7.11. Show that if $z = ke^{-r}i$, with $k \in K = SO(2)$ and $r > 0$, then

$$\log|z-i| \sim \log(\cosh r - 1) \quad \text{as} \quad r \to 0+.$$

Hint. *From Exercise 3.7.8, we have*

$$2\cosh d(z,i) + 2 = \frac{|z+i|^2}{\operatorname{Im} z}.$$

Thus,

$$2\cosh d(z,i) - 2 = \frac{|z-i|^2}{\operatorname{Im} z}.$$

Next note (see Erdélyi et al. [164, Vol. I, p. 163]) that if $s \neq -1, -2, \ldots$,

$$Q_{s-1}(\cosh r) \sim -\frac{1}{2}\log(\cosh r - 1) \qquad \text{as } r \to 0+\,. \tag{3.148}$$

Thus, in order to obtain the correct singularity, the **(free-space) Green's function** for $\Delta - \lambda$ on H must be

$$g_\lambda(z,w) = \frac{-1}{\pi} Q_{s-1}(\cosh d(z,w)),\ \lambda = s(s-1). \tag{3.149}$$

We can check this against Elstrodt [156, p. 67], where g_λ is given as follows, once we note that Elstrodt sums over Γ and not $\Gamma/\{\pm I\}$ and replaces Δ by $-\Delta$:

$$g_\lambda(z,w) = \frac{-\Gamma(s)^2}{2\pi\Gamma(2s)}\left(\frac{u+1}{2}\right)^{-s} F\left(s,s,2s;\frac{2}{u+1}\right), \tag{3.150}$$

if $u = \cosh d(z,w)$, $\lambda = s(s-1)$, and $F(a,b,c;z)$ is the Gauss hypergeometric function (see Lebedev [401, Chap. 7]). The reason is that, from Lebedev [401, p. 3 and p. 200], we have

$$\begin{aligned}\frac{-1}{\pi} Q_{s-1}(u) &= \frac{-1}{\sqrt{\pi}2^s}\frac{\Gamma(s)}{\Gamma(s+\frac{1}{2})}(u+1)^{-s}F\left(s,s,2s;\frac{2}{u+1}\right)\\ &= \frac{-1}{2\pi}\frac{\Gamma(s)^2}{\Gamma(2s)}\left(\frac{u+1}{2}\right)^{-s} F\left(s,s,2s;\frac{2}{u+1}\right).\end{aligned}$$

Next we use formula (3.147) to build up the Green's function for $\Delta - \lambda I$ on the fundamental domain $\Gamma\backslash H$.

Exercise 3.7.12. Show that the series (3.147) converges for (z,w,s) in compact sets inside the domain $H \times H \times \{s | \operatorname{Re} s > 1\}$.

Hint. *You need to know (see Lebedev [401, p. 176]) that*

$$Q_s(z) \sim \frac{\sqrt{\pi}\Gamma(s+1)}{\Gamma(s+\frac{3}{2})(2z)^{s+1}} \quad \text{as } |z| \to \infty,\ s \neq -1, -2, \ldots.$$

This allows you to bound the series (3.147) by a constant (dependent on s) times the ***non-Euclidean Eisenstein series*** *or* ***Epstein zeta function***

$$Z_\Gamma(z,w;s) = \sum_{\gamma\in\Gamma/\{\pm I\}} (\cosh d(z,\gamma w))^{-s}\,. \tag{3.151}$$

Note that Eq. (3.151) is the Mellin transform of the theta function defined in Eq. (3.139); i.e.,

$$\Gamma(s)Z_\Gamma(i,i,s) = \int_{a=0}^{\infty} a^{s-1}\theta_\Gamma(a)da. \tag{3.152}$$

You might find it hard to believe that the series (3.151) converges for $\operatorname{Re} s > 1$, since Epstein's zeta function in Sect. 1.4 given by

$$Z(I_4,s) = \sum_{(a,b,c,d)\neq 0} (a^2+b^2+c^2+d^2)^{-s}$$

only converges for $\operatorname{Re} s > 2$ (see Exercise 1.4.5 and Theorem 1.4.1 of Sect. 1.4). However, there are many fewer terms in the sum for $Z_\Gamma(z,w;s)$, as we saw in Eqs. (3.137) and (3.138). In fact, one could use Eq. (3.137) to do Exercise 3.7.11 and the fact that $Z_\Gamma(i,i;s)$ has a pole at $s = 1$ to prove Eq. (3.137) via a Tauberian theorem for Dirichlet series.

One can study the convergence of Eq. (3.151) using an integral test. For one has

$$\int_{w=u+iv\in\Gamma\backslash H} \sum_{\substack{\gamma\in\Gamma/\{\pm I\} \\ \cosh d(i,\gamma w)\geq 1}} (\cosh d(i,\gamma w))^{-s} \frac{du\,dv}{v^2}$$

$$= 2\pi \int_{\cosh r\geq 1} (\cosh r)^{-s} \sinh r\, dr,$$

which is finite if $\operatorname{Re} s > 1$.

Exercise 3.7.13. Apply the non-Euclidean Poisson summation formula of Theorem 3.7.3 to $Z_\Gamma(i,i;s)$ in formula (3.151), or Mellin transform formula (3.142).

It follows from Exercise 3.7.12 that the analytic continuation of the Green's function $G_\lambda(z,w)$ in $\lambda = s(s-1)$ to $\operatorname{Re}\, s = \frac{1}{2}$ is necessary in order to reach the spectrum of the non-Euclidean Laplacian. There are many ways to approach the analytic continuation of G_λ. We shall not go into this here, beyond giving a few more exercises. One wants this analytic continuation because it provides another proof of Theorem 3.7.1 (see formulas (2.24)–(2.26)), since one can deduce the spectral theory of Δ on $\Gamma\backslash H$ from the behavior of G_λ. The poles of G_λ correspond to the discrete spectrum of Δ, and the jump discontinuities as λ crosses the real axis correspond to the continuous spectrum. Recall here that real $\lambda = s(s-1)$ correspond to s with $\operatorname{Re}\, s = \frac{1}{2}$ or $s \in [0,1]$.

Exercise 3.7.14.

(a) Show that $G_\lambda(z,w)$ given by formula (3.147) is an L^2-function and thus has an L^2-expansion

$$-G_\lambda(z,w) = \sum_{n\geq 0} \frac{v_n(z)\overline{v_n(w)}}{\lambda-\lambda_n} + \frac{1}{4\pi i}\int_{\operatorname{Re} s=\frac{1}{2}} \frac{E_s(z)\overline{E_s(w)}}{\lambda - s(s-1)}ds.$$

Note that this gives a non-Euclidean Poisson sum formula for $G_\lambda(z,w)$.

(b) For absolute convergence, show that one should look at $G_\lambda - G_\mu$, since $(\lambda-\mu)R_\lambda R_\mu = R_\lambda - R_\mu$, if $R_\lambda = (\Delta - \lambda I)^{-1}$ =the resolvent.

Exercise 3.7.15. Since G_λ is Γ-invariant in $z = x+iy$, it has a Fourier expansion as a periodic function of $x = \operatorname{Re} z$. Show that this expansion is

$$-G_\lambda(z,w) = E_s(w)\frac{y^{1-s}}{2s-1} + \sum_{n\neq 0} F_n(w;s)\sqrt{y}K_{s-\frac{1}{2}}(2\pi|n|y)\exp(2\pi inx),$$

with

$$F_n(w;s) = \sum_{\gamma\in\Gamma_\infty\backslash\Gamma} \sqrt{\operatorname{Im}(\gamma w)}\, I_{s-\frac{1}{2}}(2\pi|n|\operatorname{Im}(\gamma w))\exp\left(2\pi in \operatorname{Re}(\gamma w)\right).$$

Here $I_s(z)$ denotes the I-Bessel function (see Lebedev [401, Sect. 5.7]). One must keep $\operatorname{Im} z$ large compared with $\operatorname{Im}(\gamma w)$, $\gamma\in\Gamma$, in this expansion.

Exercise 3.7.15 helps to explain Hejhal's work [263], discussed briefly in Sect. 3.5, showing that spurious eigenvalues $\lambda = s(s-1)$ of Δ on $\Gamma\backslash H$ correspond to zeros of Epstein's zeta function:

$$Z(W,s) = \zeta(2s)E_s(\rho) = \zeta(s)L(s,(-3/*)),$$

with

$$\rho = \exp(2\pi i/3) = (-1+i\sqrt{3})/2, \quad W = \frac{1}{\sqrt{3}}\begin{pmatrix} 2 & -1 \\ -1 & 2\end{pmatrix}.$$

For at zeros s of $E_s(\rho)$, the Green's function $G_{s(s-1)}(z,\rho)$ looks like a cusp form, by Exercise 3.7.15, except that it has a logarithmic singularity at $z=\rho$.

Hejhal [264] uses Fourier-type expansions of $G_\lambda(z,w)$ coming from certain hyperbolic matrices to show that the solutions of the congruence

$$y^2 \equiv 5 \pmod{L}$$

are such that $y/L \pmod 1$ is uniformly distributed as L approaches infinity, and that the finer properties of this distribution are ruled by the eigenvalues of Δ of $\Gamma\backslash H$. The result on distribution of solutions of quadratic congruences was first proved by Hooley [301]. The generalization to higher-degree congruences remains open. However, it is possible that extensions of Hejhal's Green's function identities to $SL(n,\mathbb{Z})$ might allow one to attack such questions, especially if one recalls the work of Dorothy Wallace [711] connecting units in higher-degree number fields and hyperbolic elements of $SL(n,\mathbb{Z})$. See Hejhal [267–271], and Fay [176] for related work.

Exercise 3.7.16 (The Heat Equation on $\Gamma\backslash H$). Given some initial heat distribution function f on $\Gamma\backslash H$, find $u = u(z,t)$ satisfying

$$\begin{aligned} &u_t = \Delta_z u,\ \Delta_z = y^2(\partial^2/\partial x^2 + \partial^2/\partial y^2),\\ &z = x+iy, \text{ and } u(z,0) = f(z), \text{ for } z \in \Gamma\backslash H,\ t > 0. \end{aligned}$$

Answer. *Using the notation (3.129) and (3.130),*

$$u(z,t) = \sum_{n\geq 0} A_n \exp(\lambda_n t) v_n(z) + \frac{1}{4\pi i}\int_{\mathrm{Re}\, s=\frac{1}{2}} A_s \exp\{s(s-1)t\} E_s(z)\, ds,$$

where $A_n = (f, v_n)$, and $A_s = (f, E_s)$.

One way to see this is to use the method of images to show that the fundamental solution for the heat equation on $\Gamma\backslash H$ is

$$\theta^{\#}(z,t) = \sum_{\gamma\in\Gamma/\{\pm I\}} g_t(\gamma z),$$

g_t = the fundamental solution for the heat equation on H itself given by formulas (3.37) and (3.38) of Sect. 3.2. Apply the non-Euclidean Poisson sum formula to see that

$$\theta^{\#}(z,t) = \sum_{n\geq 0} e^{\lambda_n t}\overline{v_n(i)} v_n(z) + \frac{1}{4\pi i}\int_{\mathrm{Re}\, s=\frac{1}{2}} e^{s(s-1)t}\overline{E_s(i)} E_s(z)\, ds$$

where $\lambda_n = s_n(s_n - 1)$.

Exercise 3.7.17 (Remarks on Convergence of Non-Euclidean Fourier Series). Suppose that $\Delta f \in L^2(\Gamma\backslash H)$. Show that the non-Euclidean Fourier "series" for f (see Theorem 3.7.1) converges uniformly and absolutely.

Answer. *Imitate the proof of part (3) of Theorem 1.3.1 of Sect. 1.3. Look at*

$$\begin{aligned} &\sum_{n\geq 0} |(f,v_n)v_n(z)| + \frac{1}{4\pi i}\int_{\mathrm{Re}\, s=\frac{1}{2}} |(f,E_s)E_s(z)|\ ds\\ &= \sum_{n\geq 0}\left|\frac{(\Delta f, v_n)v_n(z)}{s_n(s_n-1)}\right| + \frac{1}{4\pi i}\int_{\mathrm{Re}\, s=\frac{1}{2}}\left|\frac{(\Delta f, E_s)E_s(z)}{s(s-1)}\right| ds\\ &\leqq \left(\sum_{n\geq 0}\left|\frac{v_n(z)}{s_n(s_n-1)}\right|^2\right)^{\frac{1}{2}}\left(\sum_{n\geq 0}|(\Delta f, v_n)|^2\right)^{\frac{1}{2}}\\ &\quad + \left(\frac{1}{4\pi i}\int_{\mathrm{Re}\, s=\frac{1}{2}}\left|\frac{E_s(z)}{s(s-1)}\right|^2 ds\right)^{\frac{1}{2}}\left(\int_{\mathrm{Re}\, s=\frac{1}{2}}|(\Delta f, E_s)|^2\, ds\right)^{\frac{1}{2}} \end{aligned}$$

Then use the fact that Δf and G_λ = the Green's function for $\Delta - \lambda$ are both square-integrable.

There is another application of Theorem 3.7.1, which was given by Kaori Imai Ota [315] showing that there is a higher-dimensional analogue of Hecke's Theorem 3.6.1 of Sect. 3.6. We shall discuss this in more detail in Volume II ([667]). Briefly, one wants higher-dimensional Mellin transforms of Siegel modular forms for the symplectic group $Sp(n,\mathbb{Z})$. When $n=2$, Imai Ota shows that the converse theorem in Hecke theory can be proved using the Roelcke–Selberg spectral resolution of the Laplacian on $L^2(SL(2,\mathbb{Z})\backslash H)$. When n is larger than 2, this converse theorem requires a workable generalization of Theorem 3.7.1 for $SL(n,\mathbb{Z})$ – a subject to be discussed in Volume II [667].

Elstrodt et al. [157–160] and Mennicke [463,464] have used the non-Euclidean Poisson sum formula for groups such as $SL(2,\mathbb{Z}[i])$ (to be discussed in Volume II [667]) in order to derive many sorts of algebraic and analytic results. One very interesting idea of Mennicke [464] is to use Theorem 3.7.3 in order to study the eigenvalues of the Laplacian corresponding to Maass-type cusp forms, as well as to study the cusp forms themselves.

3.7.4 *Selberg's Trace Formula for* $SL(2,\mathbb{Z})$

The discussion of Selberg's trace formula should begin with a review of the part of Sect. 1.3 which gave an interpretation of Poisson's sum formula as a trace formula. However, the Selberg analogue of Poisson summation is complicated by the lack of commutativity of the groups G and Γ involved, as well as the lack of compactness of the fundamental domain. This section was greatly influenced by the work of Dorothy Wallace [711,712].

We wish to find traces of the"compact part" of the convolution operators in Eq. (3.134) and Lemma 3.7.2, defined by

$$L_g f(a) = \int_{G/K} f(b) g(b^{-1}a) db \tag{3.153}$$

viewed as integral operators on functions f in $L^2(\Gamma\backslash H)$, for K-invariant functions $g \in C_c^\infty(H)$. We can rewrite Eq. (3.153) as an integral operator on the fundamental domain as follows:

$$\begin{aligned} L_g f(a) &= \int_{\Gamma\backslash G/K} f(b) K_g(a,b) db, \\ K_g(a,b) &= \sum_{\gamma\in\overline{\Gamma}} g(b^{-1}\gamma a), \quad \overline{\Gamma} = \Gamma/\{\pm I\} = PSL(2,\mathbb{Z}). \end{aligned} \tag{3.154}$$

Note that we must use $\overline{\Gamma} = \Gamma/\{\pm I\} =$ **the projective linear group**, or the formula will be off by a factor of 2. For $\overline{\Gamma}\backslash H = \Gamma\backslash H$, since the fractional linear transformation corresponding to $-\gamma$ is the same as that for $+\gamma$, and

$$H = \bigcup_{\gamma \in \overline{\Gamma}} \gamma(D), \; D = \text{a fundamental domain for } \Gamma \backslash H,$$

is a disjoint union.

We will write

$$\begin{aligned} K_g(a,b) &= K_g(a(i),b(i)), \\ k_g(a,b) &= k_g(a(i),b(i)) = g(b^{-1}a), \end{aligned} \tag{3.155}$$

for $a,b \in G = SL(2,\mathbb{R})$. Here, as usual, $a(i)$ is the point in H to which the fractional linear transformation corresponding to $a \in G$ sends $i = \sqrt{-1}$.

In this setting, using the notation in Eqs. (3.129) and (3.130), involving v_n, s_n, and E_s, **Mercer's theorem** (see Theorem 1.3.8 of Sect. 1.3) becomes

$$\left. \begin{array}{l} K_g(a,b) = \hat{K}_g(a,b) + E_g(a,b), \text{ with} \\ \hat{K}_g(a,b) = \sum_{n\geq 0} \hat{g}(s_n) v_n(a(i)) \overline{v_n(b(i))} \\ E_g(a,b) = \frac{1}{4\pi i} \int_{\mathrm{Re}\, s = \frac{1}{2}} \hat{g}(s) E_s(a(i)) \overline{E_s(b(i))} ds. \end{array} \right\} \tag{3.156}$$

Here $\hat{g}$ denotes the **Helgason transform** of Sect. 3.2 (which is also called the Selberg transform in this context):

$$\hat{g}(s) = \int_H g(z) y^{\bar{s}-2} \, dx \, dy, \tag{3.157}$$

for K-invariant functions g. Formula (3.156) is equivalent to the non-Euclidean Poisson summation formula in Theorem 3.7.3 above.

Next define the **integral operator** $\hat{L}_g$ corresponding to $\hat{K}_g$ by

$$\hat{L}_g f(a) = \int_{\Gamma \backslash G/K} f(b) \hat{K}_g(a,b) db. \tag{3.158}$$

Our discussion of the Selberg trace formula begins by writing

$$\text{Trace } \hat{L}_g = \sum_{n\geq 0} \hat{g}(s_n) = \int_{\Gamma \backslash G/K} \hat{K}_g(a,a) \, da. \tag{3.159}$$

In order to proceed further, we must decompose $SL(2,\mathbb{Z})$ into conjugacy classes of various sorts, according to the Jordan forms (over $\mathbb{C}$) of these matrices.

Definition 3.7.2. Classification of Elements of $\boldsymbol{SL(2,\mathbb{Z}) = \Gamma}$ (cf. Sect. 3.3).

(1) **Central Elements**: $+I, -I$.

(2) **Parabolic Elements**: those having Jordan form $\pm \begin{pmatrix} 1 & a \\ 0 & 1 \end{pmatrix}$, with $a \neq 0$.

(3) **Elliptic Elements**: those with Jordan form $\begin{pmatrix} a & 0 \\ 0 & 1/a \end{pmatrix}$, with $a \notin \mathbb{R}$, $|a| = 1$.

(4) **Hyperbolic Elements**: those with Jordan form $\begin{pmatrix} a & 0 \\ 0 & 1/a \end{pmatrix}$, with $a \in \mathbb{R}, a \neq 1, -1, 0$. For a hyperbolic element γ, we define $a^2 = N\gamma$, the **norm** of the element, where a is chosen so that $|a| > 1$.

Definition 3.7.3. For $\alpha \in \Gamma$, the **conjugacy class** of α in Γ is

$$\{\alpha\} = \left\{ \gamma\alpha\gamma^{-1} \mid \gamma \in \Gamma \right\}. \tag{3.160}$$

Definition 3.7.4. For $\alpha \in \Gamma$, the **centralizer of $\boldsymbol{\alpha}$ in $\boldsymbol{\Gamma}$** is

$$\Gamma_\alpha = \{\gamma \in \Gamma \mid \gamma\alpha = \alpha\gamma\}. \tag{3.161}$$

Exercise 3.7.18 (Explicit Characterizations of the Conjugacy Classes in SL(2, $\mathbb{Z}$)).

(a) Show that the conjugacy classes of parabolic elements of $SL(2, \mathbb{Z})$ are represented by

$$\pm \begin{pmatrix} 1 & a \\ 0 & 1 \end{pmatrix}, \; a \in \mathbb{Z}.$$

(b) Show that the conjugacy classes of elliptic elements of $SL(2, \mathbb{Z})$ are represented by

$$\pm \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \pm \begin{pmatrix} 1 & -1 \\ 1 & 0 \end{pmatrix}.$$

(c) Show that the conjugacy classes of hyperbolic elements of $SL(2, \mathbb{Z})$ are represented by units ε_d in orders $\mathfrak{O}$ in real quadratic fields $\mathbb{Q}(\sqrt{d})$, with multiplicity $h_d =$ the narrow class number of the order $\mathfrak{O}$. The **narrow class number** is defined similarly to the class number in Sect. 1.4, except that the equivalence relation between two ideals $\mathfrak{a}, \mathfrak{b}$ in $\mathbb{Z}[\varepsilon_d]$ is

$$\mathfrak{a} \sim \mathfrak{b} \Leftrightarrow \mathfrak{a} = c\mathfrak{b} \quad \text{for } c \in \mathbb{Q}(\sqrt{d}) \quad \text{with norm } Nc = +1.$$

The order $\mathfrak{O}$ need not be the whole ring of integers in $\mathbb{Q}(\sqrt{d})$.

Hints. See Olga Taussky [659, 660], Shimura [589], Schoeneberg [562], and Dorothy Wallace [711, 712]. It helps to look at the points in $H \cup \mathbb{R}$ which are fixed by $\gamma \in \Gamma$. Part (c) is the most complicated. Following Taussky [loc. cit.] and Wallace [loc. cit.], you should note that if γ is a hyperbolic element of $SL(2, \mathbb{Z})$, the diagonal elements ε in the Jordan form of γ must be units in a real quadratic field K, since ε is a root of the characteristic polynomial of γ. The eigenvectors w_1, w_2 in K^2 generate an ideal $\mathbb{Z}w_1 \oplus \mathbb{Z}w_2$ in an order $\mathfrak{O}$ of K. Note that this order need not be the maximal order; i.e., the whole ring of integers of K defined in Sect. 1.4. However, one can still form the narrow ideal classes (which do not necessarily form a group). And one can relate the number of these narrow ideal classes to the ordinary class number of K (*see* Lang [390]). Sarnak [555] gives a description of part (c) of this exercise in the language of quadratic forms which goes back to Gauss.

The following exercise shows that the concept of primitive hyperbolic element of $SL(2,\mathbb{Z})$ is not the same thing as that of fundamental unit in the relevant real quadratic number field.

Exercise 3.7.19.

(a) Let γ be a hyperbolic element of $\Gamma = SL(2,\mathbb{Z})$. Show that the centralizer Γ_γ of γ in Γ is an infinite cyclic group generated by γ_0 which is called a **primitive hyperbolic element** *of* Γ.

(b) Show that $\begin{pmatrix} 2 & 9 \\ 1 & 5 \end{pmatrix}$ is a primitive hyperbolic matrix in Γ with eigenvalue

$$z^4 = (7+3\sqrt{5})/2, \text{ where } z = (1+\sqrt{5})/2.$$

Hint (see Buser [74, p. 228] or Hejhal [261, Chap. 1]). *Use the fact that* Γ *is discrete as well as the possibility of simultaneously diagonalizing commuting elements of* Γ.

Exercise 3.7.20 (Hyperbolic Conjugacy Classes and the Length Spectrum of Closed Geodesics in the Fundamental Domain).

(a) Let $C(z,w)$ be a geodesic line or circle in H connecting two points z, w on $\mathbb{R} \cup \{\infty\}$. Consider the image $\overline{C(z,w)}$ in the standard fundamental domain for $SL(2,\mathbb{Z})\backslash H$ which is given in Fig. 3.14. We say that $\overline{C(z,w)}$ is a **closed geodesic** if it is a closed curve in the fundamental domain. Here we mean "closed" in the sense that, once correctly parameterized, the beginning of the curve is the same point as the end of the curve. We do not refer to the topological notion of closed (i.e., the set of points on the curve has open complement). Show that $\overline{C(z,w)}$ is a closed geodesic in $SL(2,\mathbb{Z})\backslash H$ if and only if there is an element γ of $SL(2,\mathbb{Z})$ such that $\gamma C(z,w) \subset C(z,w)$ (as sets of points). Show then that $\overline{C(z,w)}$ is a closed geodesic if and only if z and w are the fixed points of a hyperbolic element γ of $SL(2,\mathbb{Z})$.

(b) Suppose that $\overline{C(z,w)}$ is a closed geodesic in $SL(2,\mathbb{Z})\backslash H$, with z, w the fixed points of a hyperbolic element γ of $SL(2,\mathbb{Z})$. Show that if a point q lies on $C(z,w)$, then so does γq and the hyperbolic distance between q and γq is $\log N\gamma$, where $N\gamma$ is the norm of γ, from Definition 3.7.2.

(c) Using a computer, graph $\overline{C(z,w)}$ for various choices of z, w. We did this in Figs. 3.42–3.45 using the image of the standard fundamental domain under $z \to -1/z$.

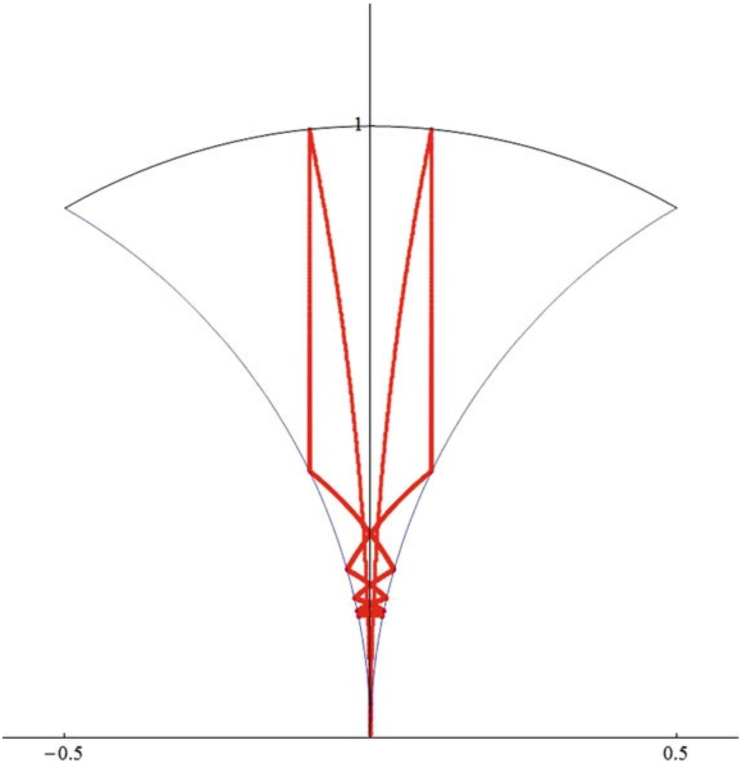


Fig. 3.42 Images of points on the geodesic circle of center 0 and radius 5 after mapping by $SL(2,\mathbb{Z})$ into the image of the standard fundamental domain under $z \to -1/z$

It follows from Exercises 3.7.18–3.7.20 that the primitive hyperbolic conjugacy classes in $SL(2,\mathbb{Z})$ have both a number-theoretic and a geometric interpretation. The number-theoretic interpretation is that they correspond to fundamental units in orders in real quadratic number fields. The geometric interpretation is that they correspond to closed or periodic geodesics on the Riemann surface $SL(2,\mathbb{Z})\backslash H$—the **length spectrum** of $SL(2,\mathbb{Z})\backslash H$. The number of hyperbolic conjugacy classes with a given trace can thus be viewed either as a class number or as the number of closed geodesics with a given length. Selberg's trace formula (Theorem 3.7.4) gives relations between the length spectrum and the eigenvalue spectrum of Δ on $SL(2,\mathbb{Z})\backslash H$. This results in an analogue of the prime number theorem (see Theorem 3.7.6), which can thus be given either a number-theoretic or a geometric interpretation. We can also view the hyperbolic conjugacy classes as corresponding to non-Euclidean billiards on $SL(2,\mathbb{Z})\backslash H$ such as that pictured in Fig. 3.43.

In Sect. 1.5.4, we considered Weyl's criterion for the equidistribution of numbers on a circle $\mathbb{R}/\mathbb{Z}$ or the torus $\mathbb{R}^2/\mathbb{Z}^2$, showing that a line in $\mathbb{R}^2$ which makes an irrational angle with the x-axis will correspond to a densely wound line in the torus

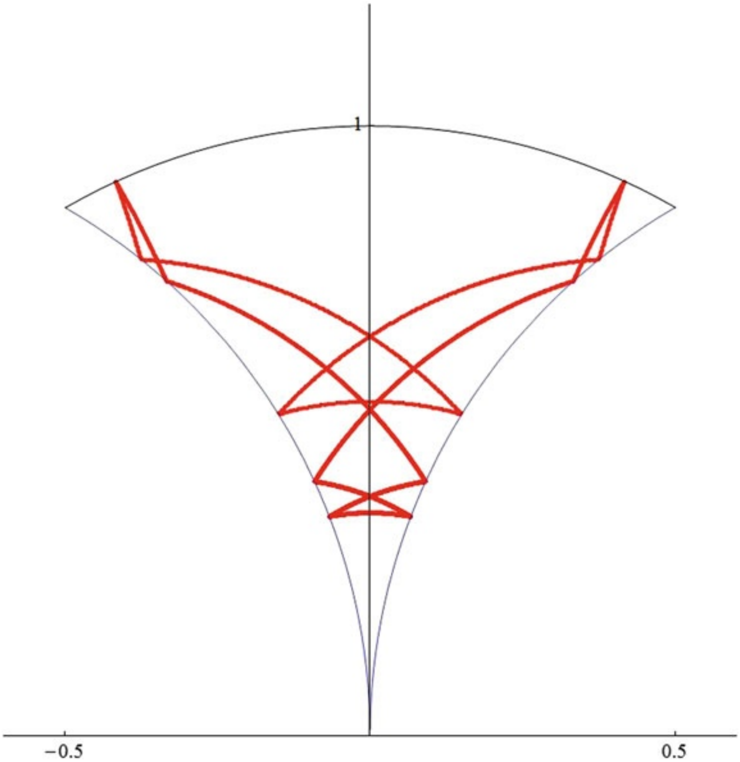


Fig. 3.43 Images of points on the geodesic circle of center 0 and radius $(15/2)^{1/2}$ after mapping by $SL(2,\mathbb{Z})$ into the image of the standard fundamental domain under $z \to -1/z$

$\mathbb{R}^2/\mathbb{Z}^2$. In Exercise 3.6.17 of Sect. 3.6, we saw that the image of a horocycle at height y in $SL(2,\mathbb{Z})\backslash H$ tends to fill up the fundamental domain as $y \to 0+$.

One can ask analogous questions about geodesics. Before asking these questions, the reader should consider Figs. 3.42–3.45, which give plots of points on a fundamental domain in various geodesic circles $\overline{C(-x,+x)}$. Figure 3.42 shows the geodesic $\overline{C(-5,+5)}$. This geodesic has only a finite number of segments and is clearly not dense, but it is also not closed. Figure 3.43 shows a genuine closed geodesic $\overline{C(-x,+x)}$ for $x = \sqrt{15/2}$, which is fixed (as is $-x$) by the hyperbolic matrix

$$\begin{pmatrix} 11 & 30 \\ 4 & 11 \end{pmatrix}.$$

Figures 3.44 and 3.45 show points on approximations to $\overline{C(-x,+x)}$ for $x = \sqrt{163}$ and $e = 3.1415\ldots$. In the first edition of this book the analogous plots were made with a TRS-80, Model 100 computer. Here I used my 2006 HP PC and Mathematica.

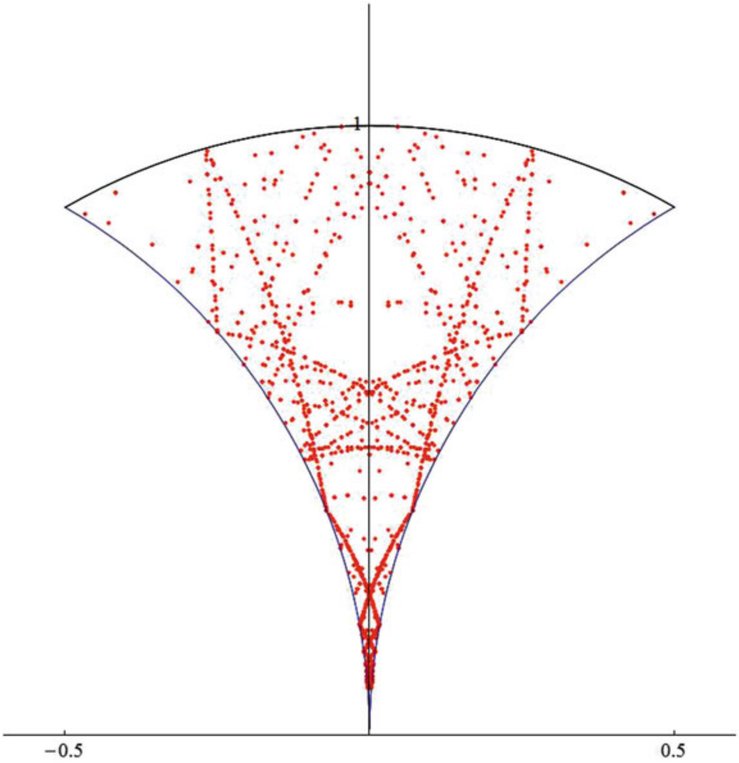


Fig. 3.44 Images of points on the geodesic circle of center 0 and radius $(163)^{1/2}$ after mapping by $SL(2,\mathbb{Z})$ into the image of the standard fundamental domain under $z \to -1/z$

Of course, the problem of drawing these pictures for geodesics with irrational radius is similar to the problem one has when drawing fractals: Computers have a fixed precision. And so does the human eye.

Questions.

(1) As the length of a closed geodesic approaches infinity, does the geodesic tend to fill up a fundamental domain, in an analogous way to that for the horocycle of Exercise 3.6.17 of Sect. 3.6?

(2) Can one give a criterion on z,w which will insure that a geodesic $C(z,w)$ is dense in a fundamental domain for $SL(2,\mathbb{Z})$?

These questions are related to ergodic theory (as well as continued fraction expansions of real numbers). Let $T : M \to M$ be a measure-preserving transformation of a Riemannian manifold. One says that T is **ergodic** if for every Lebesgue integrable function f on M, one has

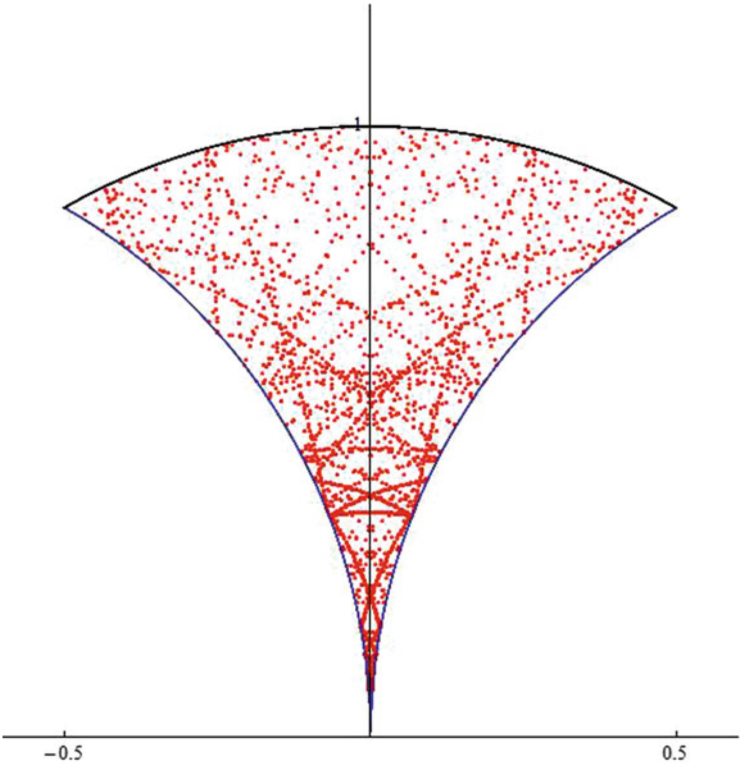


Fig. 3.45 Images of points on the geodesic circle of center 0 and radius e after mapping by $SL(2,\mathbb{Z})$ into the image of the standard fundamental domain under $z \to -1/z$

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} f(T^k x) = \int_M f(y)d\mu(y) \quad \text{for almost all } x \in M.$$

The average on the left may be considered a time average, while that on the right may be considered a space average. Actually, M is often replaced by its unit tangent bundle in most works on the subject. References for ergodic theory, geodesic flows, etc., are Artin [14], Auslander et al. [18], Bedford et al. [32], Bowen [60], Gelfand and Fomin [202], Mautner [460], Moeckel [472], and Nicholls [495]. Svetlana Katok and Ilie Ugarcovici [343] survey some of the results on coding closed geodesics in $SL(2,\mathbb{Z})\backslash H$. In particular, they give various methods to code the geodesics with bi-infinite sequences of integers. The geometric method records the sides of a given fundamental domain cut by the geodesic. The arithmetic codes use various types of continued fraction expansions of the endpoints of the geodesics or their reciprocals. If you write programs to draw your own versions of Figs. 3.42–3.45, you will appreciate the appearance of continued fractions here. Other references for the distribution of geodesics in the fundamental domain of the modular group are: Duke [144] and Einsiedler et al. [154].

To return to our discussion of the trace formula, first suppose that Γ has no parabolic elements. Then we can write the trace as follows, using formulas (3.154), (3.155), (3.159)–(3.161):

$$\begin{aligned}\sum_{n\geq 0}\hat{g}(s_n) &= \int_{\Gamma\backslash H} K_g(b,b)\,db\\ &= \sum_{\substack{\{\gamma\}\\ \text{distinct in }\overline{\Gamma}}}\ \sum_{\sigma\in\overline{\Gamma}_\gamma\backslash\overline{\Gamma}}\int_{\Gamma\backslash H} g(a^{-1}\sigma^{-1}\gamma\sigma a)\,da \qquad (3.162)\\ &= \sum_{\substack{\{\gamma\}\\ \text{distinct in }\overline{\Gamma}}}\int_{\Gamma_\gamma\backslash H} g(x^{-1}\gamma x)\,dx\end{aligned}$$

Exercise 3.7.21. *Use* the relation between K_g and k_g given in formulas (3.154) and (3.155) as well definitions (3.160) and (3.161) to prove formula (3.162) for groups Γ with compact fundamental domain.

However, in the case $\Gamma = SL(2,\mathbb{Z})$, there are parabolic elements and thus we must cancel their contribution to the trace formula against that of the continuous spectrum. This allows us to write the trace formula in the following form:

$$\sum_{n\geq 0}\hat{g}(s_n) = c_g(\infty) + \sum_{\substack{\{\gamma\}\\ \text{distinct nonparabolic}\\ \text{conjugacy classes in }\overline{\Gamma}}} c_g(\gamma), \qquad (3.163)$$

where the **orbital integrals** $c_g(\gamma)$ are defined by

$$c_g(\gamma) = \int_{\Gamma_\gamma\backslash G/K} g(a^{-1}\gamma a)\,da \quad \text{for } \gamma \quad \text{not parabolic}, \qquad (3.164)$$

and

$$c_g(\infty) = \lim_{A\to\infty}\left\{\int_{y=0}^{A}\int_{x=-\frac{1}{2}}^{\frac{1}{2}}\sum_{\substack{\gamma=\begin{pmatrix}1 & n\\ 0 & 1\end{pmatrix}\\ n\in\mathbb{Z}-0}} k_g(\gamma z,z)\frac{dx\,dy}{y^2} - \int_{\substack{z\in\Gamma\backslash H\\ y\leq A}} E_g(z,z)\frac{dx\,dy}{y^2}\right\}. \qquad (3.165)$$

Here E_g is defined by formula (3.156) and $\Gamma\backslash H$ means the standard fundamental domain for $SL(2,\mathbb{Z}) = \Gamma$ (as in Fig. 3.14).

In order to put the trace formula in final form, Selberg showed how to evaluate the various $c(\gamma)$ above in terms of $\hat{g}$ or a closely related integral—the **Harish transform** (or Abel transform or horocycle transform) defined by

$$Tg(y) = y^{-\frac{1}{2}} \int_{x\in\mathbb{R}} g(x+iy)dx = G(\log y). \tag{3.166}$$

We know from Exercise 3.2.21 of Sect. 3.2 that

$$\hat{g}(s) = MTg\left(\overline{s} - \frac{1}{2}\right), \tag{3.167}$$

where M denotes the Mellin transform.

Now we are ready to prove Selberg's trace formula for $SL(2,\mathbb{Z})$. It should be compared with the result of Exercise 1.3.11 which gives a trace formula on $\mathbb{R}^2/\mathbb{Z}^2$ and with the finite analogue stated in Sect. 3.7.8.

Theorem 3.7.4 (Selberg's Trace Formula). *Suppose that $g : H \to \mathbb{C}$ is in $C_c^\infty(K\backslash G/K)$; i.e., g is a compactly supported infinitely differentiable K-invariant function on H. Then,using the notation (3.157), (3.160), and (3.166), we have*

$$\begin{aligned}\sum_{n\geq 0} \hat{g}(s_n) = {} & \frac{\text{area}(\Gamma\backslash H)}{4\pi} \int_{r\in\mathbb{R}} \hat{g}\left(\frac{1}{2}+ir\right) r\tanh(\pi r)\ dr \\ & + \sum_{\substack{\{\gamma_0\} \\ \text{primitive} \\ \text{hyperbolic}}} \sum_{k\geq 1} \frac{\log N\gamma_0}{N\gamma_0^{k/2} - N\gamma_0^{-k/2}} G(k\log N\gamma_0) \\ & + \int_{r\in\mathbb{R}} \left(\frac{1}{4} + \frac{e^{\pi r/3}+e^{-\pi r/3}}{3\sqrt{3}}\right) \frac{\hat{g}(\frac{1}{2}+ir)}{e^{\pi r}+e^{-nr}} dr \\ & - G(0)\log(2\pi) + \frac{1}{\pi}\int_{r\in\mathbb{R}} \hat{g}\left(\frac{1}{2}+ir\right) \frac{\zeta'}{\zeta}(-2ir)\ dr.\end{aligned}$$

Proof. We begin with formulas (3.163)–(3.165). It remains to evaluate the various orbital integrals in terms of the Helgason and Harish transforms of g.

Central Term. We seek to prove

$$c_g(I) = \frac{\text{area}(\Gamma\backslash H)}{4\pi} \int_{r\in\mathbb{R}} \hat{g}(\frac{1}{2}+ir)\ r\tanh(\pi r)\, dr. \tag{3.168}$$

It follows immediately from the definitions that

$$c_g(I) = \int_{\Gamma\backslash H} g(a^{-1}Ia)da = g(I)\ \text{area}(\Gamma\backslash H).$$

The inversion formula for the Helgason transform of a K-invariant function [formula (3.34) completes the proof of formula (3.168)].

Hyperbolic Term. Suppose that $\gamma = \gamma_0^k$, where γ_0 = a primitive hyperbolic in Γ. Then we want to prove

$$c_g(\gamma) = \frac{\log N\gamma_0}{N\gamma_0^{k/2} - N\gamma_0^{-k/2}} G(k \log N\gamma_0). \tag{3.169}$$

Use Exercise 3.7.19 to see that if γ is hyperbolic then

$$\Gamma_\gamma = \{\, \gamma_0^k \;\big|\; k \in \mathbb{Z}\} \quad \text{with } \gamma_0 = \text{primitive hyperbolic.}$$

Then replace γ by $\xi\gamma\xi^{-1}$, $\xi \in G$, with $\xi\gamma\xi^{-1}$ diagonal. Also replace Γ by $\xi\Gamma\xi^{-1}$.

Thus, in our computation of $c_g(\gamma)$, we can take

$$\Gamma_\gamma = \{\, \gamma_0^k \;\big|\; k \in \mathbb{Z}\},\ \gamma_0 = \begin{pmatrix} a & 0 \\ 0 & 1/a \end{pmatrix}, \quad a^2 = N\gamma_0 \geq 1. \tag{3.170}$$

We will need the following exercise.

Exercise 3.7.22. Show that if Γ_γ is given by Eq. (3.170), then we can choose a fundamental domain for $\Gamma_\gamma \backslash H$ to be

$$\{z \in H \mid 1 \leq y \leq N\gamma_0\}.$$

As usual, we identify z in H with $a_{x,y} \in SL(2,\mathbb{R})$ via

$$z = x + iy \in H \to a_{x,y} = \begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \sqrt{y} & 0 \\ 0 & 1/\sqrt{y} \end{pmatrix} \in SL(2,\mathbb{R}). \tag{3.171}$$

Therefore, if $\gamma = \gamma_0^k$, $\gamma_0 = \begin{pmatrix} a & 0 \\ 0 & 1/a \end{pmatrix}$, with $a \geq 1$, we have

$$c_g(\gamma) = \int_{\Gamma_\gamma \backslash H} g(a_{x,y}^{-1} \gamma a_{x,y}) y^{-2} dx\, dy.$$

Next note that

$$a_{x,y}^{-1} \gamma a_{x,y} = \begin{pmatrix} 1 & -x/y \\ 0 & 1 \end{pmatrix} \begin{pmatrix} a^k & 0 \\ 0 & a^{-k} \end{pmatrix} \begin{pmatrix} 1 & x/y \\ 0 & 1 \end{pmatrix}.$$

So we can make the change of variables $x \to u = x/y,\ du = dx/y$ to obtain

$$
\begin{aligned}
c_g(\gamma) &= \int_{u\in\mathbb{R}}\int_{y=1}^{a^2=N\gamma_0} g\left(\begin{pmatrix}1 & -u\\ 0 & 1\end{pmatrix}\begin{pmatrix}a^k & 0\\ 0 & a^{-k}\end{pmatrix}\begin{pmatrix}1 & u\\ 0 & 1\end{pmatrix}\right)\, du\, y^{-1}dy\\
&= \log N\gamma_0 \int_{u\in\mathbb{R}} g\left(a^{2k}(i+u)-u\right)\, du.
\end{aligned}
$$

Set $w=(a^{2k}-1)u$ to find that formula (3.169) is indeed correct.

Elliptic Term. If γ_θ is an elliptic element of Γ with $\mathrm{Tr}\gamma_\theta = 2\cos\theta$, then $\theta=\pi/2$ or $\pi/3$. Let m_θ be the order of the centralizer $\overline{\Gamma}\gamma_\theta = \Gamma\gamma_\theta/\{\pm 1\}$. Then $m_{\pi/2}=2$ and $m_{\pi/3}=3$. We shall prove that

$$
c_g(\gamma_\theta) = \frac{1}{4m_\theta\sin\theta}\int_{r\in\mathbb{R}} \frac{\cosh[(\pi-2\theta)r]}{\cosh(\pi r)}\hat{g}\left(\frac{1}{2}+ir\right)\, dr. \tag{3.172}
$$

Discussions of this term appear in Hejhal [261, Vol. I, pp. 351 ff] and Kubota [375, pp. 99 ff]. Lang [389, pp. 166–167] gives relations between elliptic and hyperbolic orbital integrals, which may be of use in simplifying the proof.

Upon replacing γ_θ by a conjugate, we can assume that

$$
\gamma_\theta = \begin{pmatrix}\cos\theta & \sin\theta\\ -\sin\theta & \cos\theta\end{pmatrix} \in K.
$$

Then we use geodesic polar coordinates from Sect. 3.1 on the orbital integral, plus the fact that $K=SO(2)$ is abelian, to see that (if $m=m_\theta$)

$$
\begin{aligned}
c_g(\gamma_\theta) &= \frac{1}{m}\int_{G/K} g(x^{-1}\gamma_\theta x)\, dx\\
&= \frac{2\pi}{m}\int_{v>0} g(a_{-v}\gamma_\theta a_v)\sinh v\, dv,\ a_v = \begin{pmatrix}e^{-v/2} & 0\\ 0 & e^{v/2}\end{pmatrix}.
\end{aligned}
$$

Next we want to make use of Fourier inversion on G/K [see formulas (3.32)–(3.34)] which says that

$$
g(ke^{-p}i) = \frac{1}{4\pi}\int_{r\in\mathbb{R}} \hat{g}(\frac{1}{2}+ir)P_{-\frac{1}{2}+ir}(\cosh p)\, r\tanh(\pi r)\, dr.
$$

In order to plug this into the elliptic term and evaluate the result, we must remark that if $a_{-v}\gamma_\theta a_v = k_1 a_{v(\theta)}k_2$, with $k_i\in K$ and $a_v,\ a_{v(\theta)}$ as above, then

$$
\cosh v(\theta) = \left(1-2\sin^2\theta\right)+2\sin^2\theta\cosh^2 v. \tag{3.173}
$$

It follows that the proof of Eq. (3.172) will be completed by the following exercise.

Exercise 3.7.23. Show that if $v(\theta)$ is defined by Eq. (3.173), then

$$\int_{v>0} P_{-\frac{1}{2}+ir}(\cosh v(\theta))\sinh v\, dv = \frac{\cosh[r(\pi-2\theta)]}{2|\sin\theta|r\sinh(\pi r)}.$$

Hints *(see Sneddon [609, pp. 384–387] or Erdélyi et al. [165, Vol. II, p. 330]). Make the change of variables from v to* $w = \cosh v(\theta)$*. Then you must compute the Mehler–Fock transform*

$$\int_{w\geq 1} (w-\cos 2\theta)^{-1/2} P_{-\frac{1}{2}+ir}(w)\, dw, \quad \text{assuming } \cos(2\theta)\geq 0.$$

To compute such a Mehler–Fock transform, you can use the following formulas.

(a)

$$\int_0^\infty g(t)P_{-\frac{1}{2}+it}(\cosh a)\tanh(\pi t)\ dt = \pi^{-1/2}\int_a^\infty (\cosh t-\cosh a)^{-1/2} Fg(t)dt,$$

where

$$Fg(t) = \sqrt{\frac{\pi}{2}}\int_0^\infty g(v)\sin(tv)dv.$$

(b) When $g(v) = \cosh(bv)/\sinh(\pi v)$, $-\pi < b < \pi$, *we have*

$$Fg(t) = \frac{1}{\sqrt{2\pi}}\frac{\sinh t}{\cosh t+\cos b}.$$

(c)

$$\int_a^\infty \frac{\sinh t\ dt}{(\cosh t+\cos b)\sqrt{\cosh t-\cosh a}} = \frac{\pi}{\sqrt{\cosh a+\cos b}}.$$

Parabolic Term. We must show that

$$c_g(\infty) = -G(0)\log(2\pi) + \frac{1}{\pi}\int_{r\in\mathbb{R}} \hat{g}\left(\frac{1}{2}+ir\right)\frac{\zeta'}{\zeta}(-2ir)\, dr. \tag{3.174}$$

Begin by noting that conjugacy classes of parabolic elements of $PSL(2,\mathbb{Z})$ are represented by $\gamma_n(z) = z+n$, $n\in\mathbb{Z}$, with centralizer

$$\Gamma_{\gamma_n} = \{\gamma_m \mid m\in\mathbb{Z}\}.$$

Clearly

$$\Gamma_{\gamma_n}\backslash H = \left\{ x+iy \;\middle|\; |x| \le \frac{1}{2},\ y>0 \right\}.$$

Then we define, as in Eq. (3.165), $c_g(\infty) = \lim_{A\to\infty}\{T_1(A) - T_2(A)\}$, where

$$\begin{aligned} T_1(A) &= \int_{y=0}^{A}\int_{x=-\frac{1}{2}}^{\frac{1}{2}} \sum_{n\neq 0} g\left(a_{x,y}^{-1} \begin{pmatrix} 1 & n \\ 0 & 1 \end{pmatrix} a_{x,y} \right) \frac{dx\,dy}{y^2}, \\ T_2(A) &= \frac{1}{4\pi i} \int\limits_{\substack{\mathrm{Re}\, s=\frac{1}{2} \\ x^2+y^2\ge 1 \\ |x|\le\frac{1}{2},\ y<A}} \hat{g}(s) E_s(z) E_{\bar{s}}(z)\, ds\, \frac{dx\,dy}{y^2}, \\ a_{x,y} &= \begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix} \begin{pmatrix} y^{\frac{1}{2}} & 0 \\ 0 & y^{-\frac{1}{2}} \end{pmatrix}. \end{aligned} \tag{3.175}$$

We shall follow Selberg's Göttingen lectures [573, p. 79]. This method should be compared with that of Kubota [375, p. 102 ff]. One should also consider the method of Warner [718]. See the article of Gelbart and Jacquet in the Corvallis conference proceedings (Borel and Casselman [52, Vol. I, p. 245]) for an expression for a parabolic (or nilpotent) orbital integral as a limit of hyperbolic orbital integrals. Barbasch [23, 24] pursues this matter for general Lie groups.

Let us start with $T_1(A)$. Note that

$$a_{x,y}^{-1} \begin{pmatrix} 1 & n \\ 0 & 1 \end{pmatrix} a_{x,y} = \begin{pmatrix} y^{-\frac{1}{2}} & 0 \\ 0 & y^{\frac{1}{2}} \end{pmatrix} \begin{pmatrix} 1 & n \\ 0 & 1 \end{pmatrix} \begin{pmatrix} y^{\frac{1}{2}} & 0 \\ 0 & y^{-\frac{1}{2}} \end{pmatrix} = \begin{pmatrix} 1 & n/y \\ 0 & 1 \end{pmatrix}.$$

Thus

$$T_1(A) = \int\limits_{y=0}^{A} \sum_{n\neq 0} g\left(i + \frac{n}{y}\right) \frac{dy}{y^2} = \int\limits_{w=1/A}^{\infty} \sum_{n\neq 0} g(i+nw)\,dw,$$

setting $w = 1/y$. It follows that

$$T_1(A) = 2\int_0^\infty g(i+u) \sum_{n=1}^{Au} \frac{1}{n}\, du. \tag{3.176}$$

Formula (3.176) is a reasonable one itself, but we want to express everything in terms of G or $\hat{g}$. So we make use of the following exercise.

Exercise 3.7.24. Show that

$$\sum_{0<n\le x} \frac{1}{n} = \log x + \gamma + \mathcal{O}\left(\frac{1}{x}\right) \quad \text{as } x\to\infty,$$

where γ =Euler's constant $= 0.57721566\ldots$.

Hint. *This can be proved using the Euler–Maclaurin summation formula (see Edwards [149, p. 285]).*

It follows from Exercise 3.7.24 and formula (3.176) that

$$T_1(A) = (\log A + \gamma)G(0) + 2\int_0^\infty \log u \;\; g(i+u)\,du + \mathrm{o}(1) \quad \text{as } A \to \infty. \tag{3.177}$$

To complete the evaluation of $T_1(A)$, we need another exercise.

Exercise 3.7.25. Show that

$$\begin{aligned} 2\int_0^\infty \log u\, g(i+u)\,du = &-(\log 2+\gamma)\,G(0) + \frac{1}{4}\hat{g}(\frac{1}{2}) \\ &- \frac{1}{2\pi}\int_{r\in\mathbb{R}} \hat{g}\left(\frac{1}{2}+ir\right)\frac{\Gamma'}{\Gamma}(1-ir)\,dr. \end{aligned}$$

Hint *(See Selberg [573]).*

Step 1. *Let $g(a_{x,y}) = h_g\left(\mathrm{Tr}({}^t a_{x,y} a_{x,y}) - 2\right)$, if $a_{x,y}$ is as in formula (3.175); i.e.,*

$$g(x+iy) = h_g\left(y - 2 + \frac{1+x^2}{y}\right).$$

Next let

$$H_g(v) = \int_{-\infty}^\infty h_g(v+u^2/2)\,du.$$

Then by Exercise 3.2.21 of Sect. 3.2, we have

$$h_g(v) = \frac{-1}{2\pi}\int_{w\in\mathbb{R}} H_g'(v+w^2/2)\,dw.$$

We can relate H_g and the Harish transform (3.166) of g as in Exercise 3.2.21 of Sect. 3.2:

$$\frac{1}{\sqrt{2}}H_g(y+y^{-1}-2) = G(\log y) = \frac{1}{\sqrt{y}}\int_{x\in\mathbb{R}} h_g\left(y-2+\frac{1+x^2}{y}\right)\,dx.$$

Step 2. *So we must deal with*

$$2\int_0^\infty \log u\, h_g(u^2)\,du = 2\int_0^\infty \log u\left(\frac{-1}{2\pi}\int_{w\in\mathbb{R}} H_g'\left(u^2+\frac{w^2}{2}\right)dw\right)du.$$

Set $u^2+w^2/2 = v$, and the integral becomes

$$-\frac{2}{\pi\sqrt{2}}\int_0^\infty \log u \int_{v\geq u^2} H_g'(v)\frac{dv\,du}{\sqrt{v-u^2}}.$$

Next reverse the order of integration and obtain

$$-\frac{2}{\pi\sqrt{2}}\int_{v=0}^\infty H_g'(v)\int_{u=0}^{\sqrt{v}} \log u\,\frac{du\,dv}{\sqrt{v-u^2}}.$$

Then set $u = x\sqrt{v}$ to turn the integral into

$$\begin{aligned}&-\frac{2}{\pi\sqrt{2}}\int_{v=0}^\infty H_g'(v)\int_{x=0}^{1} (\log\sqrt{v}+\log x)\frac{dx\,dv}{\sqrt{1-x^2}}\\ &= -\frac{1}{2\sqrt{2}}\int_{v=0}^\infty H_g'(v)\log v\,dv - \log 2\,G(0).\end{aligned}$$

For the last equality, you can use Gradshteyn and Ryzhik [228, formula (3.452) on p. 335].

Step 3. *Note that*

$$\begin{aligned}\int_0^\infty \log u\,H_g'(u)du &= \int_0^\infty \log(e^x+e^{-x}-2)\sqrt{2}\,G'(x)\,dx\\ &= \sqrt{2}\int_0^\infty xG'(x)dx + 2\sqrt{2}\int_0^\infty \log(1-e^{-x})G'(x)\,dx\\ &= -\sqrt{2}\int_{x=0}^\infty G(x)dx\; + 2\sqrt{2}\int_{x=0}^\infty \log(1-e^{-x})G'(x)\,dx\\ &= \frac{-1}{\sqrt{2}}\hat{g}\left(\frac{1}{2}\right) + 2^{3/2}\int_0^\infty \log(1-e^{-x})G'(x)\,dx.\end{aligned}$$

Step 4.

$$\int_{x=0}^\infty \log(1-e^{-x})G'(x)dx = \gamma G(0) + \frac{1}{2\pi}\int_{r\in\mathbb{R}} \hat{g}\left(\frac{1}{2}+ir\right)\frac{\Gamma'}{\Gamma}(1-ir)\,dr.$$

To see this, use

$$G(x) = \frac{1}{2\pi}\int_{r\in\mathbb{R}} e^{ixr}\hat{g}\left(\frac{1}{2}+ir\right)\,dr$$

and

$$ir\int_{x=0}^\infty e^{ixr}\log(1-e^{-x})dx = \frac{\Gamma'}{\Gamma}(1-ir)+\gamma,$$

from Erdélyi et al. [165, Vol. I, p. 316]. This will complete Exercise 3.7.25.

Finally we can use formula (3.177) and Exercise 3.7.25 to obtain

$$T_1(A)=G(0)\log\frac{A}{2}+\frac{1}{4}\hat{g}\left(\frac{1}{2}\right)-\frac{1}{2\pi}\int_{r\in\mathbb{R}}\hat{g}\left(\frac{1}{2}+ir\right)\frac{\Gamma'}{\Gamma}(1-ir)\,dr+o(1),\text{ as }A\to\infty. \tag{3.178}$$

Now we turn to the second term in $c_g(\infty)$ and use Exercise 3.5.12 of Sect. 3.5 to show that as $A\to\infty$, we have

$$T_2(A)=\frac{1}{4\pi}\int\limits_{\substack{r\in\mathbb{R}\\ s=\frac{1}{2}+ir}}\hat{g}(s)\left(2\log A-\frac{\varphi'}{\varphi}(s)+\frac{\varphi(\overline{s})A^{2ir}-\varphi(s)A^{-2ir}}{2ir}\right)dr+o(1), \tag{3.179}$$

where

$$\varphi(s)=\Lambda(1-s)/\Lambda(s)\quad\text{and }\Lambda(s)=\pi^{-s}\Gamma(s)\zeta(2s).$$

To evaluate this, we need the following exercise.

Exercise 3.7.26. Show that

$$\frac{1}{4\pi}\int\limits_{\substack{r\in\mathbb{R}\\ s=\frac{1}{2}+ir}}\hat{g}(s)\left(\frac{\varphi(\overline{s})A^{2ir}-\varphi(s)A^{-2ir}}{2ir}\right)dr=\frac{1}{4}\hat{g}\left(\frac{1}{2}\right)\varphi\left(\frac{1}{2}\right)+o(1),\text{ as }A\to\infty.$$

Note that $\varphi\left(\frac{1}{2}\right)=-1$.

Hints. You could use the residue theorem as in Selberg [573], or beginning facts about Fourier integrals (see Sect. 1.2 and Kubota [375]).

It follows from Eq. (3.179) and Exercise 3.7.26 that

$$\begin{aligned}T_2(A)&=\frac{\log A}{2\pi}\int_{r\in\mathbb{R}}\hat{g}\left(\frac{1}{2}+ir\right)dr-\frac{1}{4\pi}\int\limits_{\substack{r\in\mathbb{R}\\ s=\frac{1}{2}+ir}}\hat{g}(s)\frac{\varphi'}{\varphi}(s)dr+\frac{1}{4}\hat{g}\left(\frac{1}{2}\right)\\&\quad+o(1),\ \text{ as }\ A\to\infty,\\&=\log\,A\,G(0)-\frac{1}{4\pi}\int_{r\in\mathbb{R}}\hat{g}\left(\frac{1}{2}+ir\right)\frac{\varphi'}{\varphi}\left(\frac{1}{2}+ir\right)dr+\frac{1}{4}\hat{g}\left(\frac{1}{2}\right)\\&\quad+o(1),\ \text{ as }\ A\to\infty.\end{aligned}$$

So this combines with Eqs. (3.178) and (3.165) plus (3.175) to give

$$c_g(\infty)=-G(0)\log 2-\frac{1}{2\pi}\int_{r\in\mathbb{R}}\hat{g}\left(\frac{1}{2}+ir\right)\left(\frac{\Gamma'}{\Gamma}(1-ir)-\frac{1}{2}\frac{\varphi'}{\varphi}\left(\frac{1}{2}+ir\right)\right)dr. \tag{3.180}$$

Marie-France Vignéras [700] has simplified formula (3.180) as in the following exercise.

Exercise 3.7.27. Use the functional equation of Riemann's zeta function to show that

$$c_g(\infty) = -G(0)\log 2\pi + \frac{1}{\pi}\int_{r\in\mathbb{R}} \hat{g}\left(\frac{1}{2}+ir\right)\frac{\zeta'}{\zeta}(-2ir)\,dr.$$

This completes the proof of formula (3.174) for the parabolic term. Combining formulas (3.168), (3.169), (3.172), and (3.174) gives the formula in Theorem 3.7.4 and completes the proof of Selberg's trace formula, once one has done Exercise 3.7.29. □

Our discussion of the parabolic term of the trace formula was very indirect. It would be nice to find a simplification. In particular, all the cancellations that occurred during the computation of the parabolic term make one suspect that there is a more direct route to the result given in Exercises 3.7.24–3.7.27.

Exercise 3.7.28. Show that

$$\left|\frac{\varphi'}{\varphi}(s)\right| \leq C\log(2+|r|) \quad \text{for } s=\frac{1}{2}+ir,\ r\in\mathbb{R}.$$

Hint. *See Titchmarsh [680, pp. 50–53].*

Exercise 3.7.29. Show that $|\hat{g}\left(\frac{1}{2}+ir\right)| \leq C(1+|r|)^{-2-\varepsilon}$, for $\varepsilon > 0$, suffices to make the Selberg trace formula valid.

Selberg's trace formula is surprisingly similar to Weil's explicit formulas relating sums over zeros of Riemann's zeta function to sums over the primes (as well as analogues for Hecke L-functions of algebraic number fields). See Cartier and Hejhal [78], Goldfeld [217], Lang [388], Murty [488], and Weil [729, Vol. II, pp. 48–61] for discussions of Weil's result, which was used by Weil to obtain a positivity condition equivalent to the Riemann hypothesis for Hecke L-functions of number fields.

3.7.5 Applications of Selberg's Trace Formula

Our first application of the trace formula is a non-Euclidean analogue of Weyl's result on the asymptotic distribution of the eigenvalues of the Laplace operator (see Theorem 1.3.5 of Sect. 1.3).

Theorem 3.7.5 (Asymptotics of the Eigenvalues of Δ on $L^2(\Gamma\backslash H)$). *Set*

$$N(x) = \sum_{|\lambda|\leq x} \dim \mathcal{SN}(SL(2,\mathbb{Z}),\lambda),$$

where $\mathcal{SN}(\Gamma,\lambda)$ is the space of Maass waveforms which are cusp forms, as defined in Sect. 3.5. Then

$$N(x) \sim \frac{\text{area}(\Gamma\backslash H)}{4\pi} x \quad \text{as} \quad x \to \infty.$$

Proof. We use the same sort of argument that worked in the Euclidean case (see Theorem 1.3.5 of Sect. 1.3). Substitute

$$\hat{g}_t(s) = e^{s(s-1)t}$$

in Selberg's trace formula. Then $g_t(z)$ is the fundamental solution of the non-Euclidean heat equation on H (see formulas (3.37) and (3.38) and note that we have changed the notation slightly).

The left-hand side of the Selberg trace formula in Theorem 3.7.4 for $g = g_t$ is

$$\tilde{\theta}_\Gamma(t) = \sum_{n\geq 0} \exp(\lambda_n t),$$

where λ_n runs through the discrete spectrum of Δ on $SL(2,\mathbb{Z})\backslash H$. Like Eq. (3.139) the function $\tilde{\theta}_\Gamma$ can be considered to be a non-Euclidean analogue of the theta function, as we mentioned in the introduction to this section (see Molchanov [473] and Exercise 3.7.16).

Exercise 3.7.30. Show that the Harish transform of the fundamental solution of the heat equation above is

$$G_t(x) = \frac{1}{2\pi}\int_{r\in\mathbb{R}} e^{irx}\, \hat{g}_t\left(\frac{1}{2}+ir\right) dr = \frac{1}{\sqrt{4\pi t}} \exp\left\{-\left(\frac{t}{4}+\frac{x^2}{4t}\right)\right\}.$$

Next we look as **the various terms of Selberg's trace formula** *(Theorem 3.7.4) with* $g = g_t$, as $t \to 0+$.

Identity Term. Since $\tanh(\pi r) \sim 1$ as $r \to \infty$, we have

$$\frac{\text{area}(\Gamma\backslash H)}{2\pi}\int_{r=0}^{\infty} \exp\left\{-\left(\frac{1}{4}+r^2\right)t\right\} r\tanh(\pi r)\ dr \sim \frac{\text{area}(\Gamma\backslash H)}{4\pi t}, \quad t \to 0+.$$

Elliptic Term.

$$\int_{r\in\mathbb{R}}\left(\frac{1}{4}+3^{-3/2}\left(e^{\pi r/3}+e^{-\pi r/3}\right)\right)\frac{e^{-(\frac{1}{4}+r^2)t}}{e^{\pi r}+e^{-\pi r}} dr \leq \frac{1}{2}e^{-t/4}\int_0^\infty \left(\frac{1}{4}+3^{-3/2}\right)e^{-r^2 t} dr$$
$$= \mathcal{O}(t^{-\frac{1}{2}}), \qquad t \to 0+.$$

This term is thus negligible compared with the identity term.

Hyperbolic Term.

$$\sum_{\{\gamma\}\text{ hyperbolic}} \frac{\log N\gamma_0}{N\gamma^{1/2}-N\gamma^{-1/2}} \frac{e^{-t/4}}{\sqrt{4\pi t}} \exp\{-(\log N\gamma)^2/4t\} = \mathrm{o}(1),\ t \to 0+ .$$

Exercise 3.7.31. Fill in the details for the stated result on the hyperbolic term.

Parabolic Terms. Note that $G_t(0) = (4\pi t)^{-1/2} e^{-t/4}$ is negligible compared with the identity term. Then use the following exercise to see that the story is the same for the rest of the parabolic term.

Exercise 3.7.32. Show that

$$\frac{\zeta'}{\zeta}(-2ir) = \mathcal{O}(\log|r|) \quad \text{as} \quad |r| \to \infty.$$

Hint. *Use the functional equation of* ζ *to change* $-2ir$ *to* $1+2ir$. *Then use the Euler–Maclaurin summation formula, for example, to bound* $\zeta(1+2ir)$ *(see Edwards [149, p. 183]) and Stirling's formula to bound gamma.*

It follows from the Selberg trace formula and Exercises 3.7.30–3.7.32 that

$$\tilde{\theta}_r(t) \sim \text{area}\ (SL(2,\mathbb{Z})\backslash H)/(4\pi t) \quad \text{as} \quad t \to 0+ .$$

Theorem 3.7.5 is a consequence of the Tauberian theorem given as Theorem 1.2.5 of Sect. 1.2. □

It follows from Theorem 3.7.5 that the dimensions of the spaces $\mathcal{SN}(\Gamma, s(s-1)$ cannot grow too rapidly with the imaginary part of s. This answers a question that had worried us in the discussion following Theorem 3.5.4 of Sect. 3.5. Of course, it still remains open whether the dimensions can be larger than one.

It is also possible to work out error terms for the asymptotic result in Theorem 3.7.5 (see, e.g., Hejhal [261]). These applications remind one of the work of analytic number theorists on the distribution of primes (see Davenport [120], for example), using the relationship between primes and zeros of the Riemann zeta function derived from the Euler product for $\zeta(s)$ and the Weierstrass factorization of $\frac{1}{2}s(s-1)\pi^{-s/2}\Gamma(s/2)\zeta(s)$ as a product over the nonreal zeros of $\zeta(s)$. This sort of analogy motivates much of the work in Hejhal [261, 262]. In this trace-formula analogue of analytic number theory, primes are replaced by norms of the primitive hyperbolic conjugacy classes and the zeros of $\zeta(s)$ are replaced with s such that $s(s-1)$ is an eigenvalue of Δ from the discrete spectrum. Next we want to reverse the direction of the flow of information from that in Theorem 3.7.5 and use the trace formula to study the $N\gamma$, $\{\gamma\}$ hyperbolic.

Theorem 3.7.6 (Sarnak's Theorem on the Asymptotics of Units in Real Quadratic Fields à la Wallace). *Let ε_d, h_d be as in Exercise 3.7.18, part (c); i.e., the ε_d are fundamental units in certain orders $\mathfrak{O}$ (not necessarily maximal) in real quadratic fields $\mathbb{Q}(\sqrt{d})$ and h_d is the narrow class number of the order $\mathfrak{O}$. Then*

$$\sum_{\varepsilon_d \leq x} h_d \sim \frac{x^2}{\log(x^2)} \quad \text{as } x \to \infty.$$

Proof. We take our discussion from Dorothy Wallace [711]. It should be compared with that in Hejhal [261, Vol. II, p. 519], Sarnak [555, 557] and Subia [647]—each of whom chooses a different function to plug into Selberg's trace formula. We choose instead the function of formula (3.141):

$$f_a(z) = \exp[-a \cosh d(i,z)] = \exp\left[-\frac{1}{2} a \operatorname{Tr}(W_z)\right],$$

for $z \in H$, and W_z the corresponding positive matrix in $\mathcal{SP}_2$ from formula (3.140). Then we have [by formula (3.141)] the Helgason transform

$$\widehat{f_a}(s) = 2(2\pi/a)^{1/2} K_{\bar{s}-\frac{1}{2}}(a).$$

And the Harish transform is

$$\begin{aligned} F_a(\log y) &= \frac{1}{\sqrt{y}} \int_{x \in \mathbb{R}} \exp\left[-\frac{a}{2}\left(\frac{1}{y} + \frac{x^2+y^2}{y}\right)\right] dx \\ &= \sqrt{\frac{2\pi}{a}} \exp\left[-\frac{a}{2}\left(y + \frac{1}{y}\right)\right]. \end{aligned}$$

Now plug this into the Selberg trace formula and use Eq. (3.143) to obtain

$$\begin{aligned} &2\sqrt{\frac{2\pi}{a}} \sum_{n \geq 0} K_{\bar{s}_n - \frac{1}{2}}(a) \\ &\sim \sqrt{\frac{2\pi}{a}} \sum_{\substack{\{\gamma\} \\ \text{hyperbolic}}} \frac{\log N\gamma_0}{N\gamma^{1/2} - N\gamma^{-1/2}} \exp[-a(N\gamma + N\gamma^{-1})], \quad a \to 0+. \end{aligned} \tag{3.181}$$

Then, we have

$$\sqrt{\frac{\pi}{a}} \sim \sum_{\substack{\gamma \\ \text{hyperbolic}}} \frac{\log N\gamma_0}{N\gamma^{1/2} - N\gamma^{-1/2}} \exp\left[-\frac{a}{2}(N\gamma + N\gamma^{-1})\right], \; a \to 0+.$$

It follows from Exercise 3.7.18, since the terms from the primitive hyperbolics dominate the rest, that

$$\sqrt{\frac{\pi}{a}} \sim \sum_{\varepsilon_d} h_d \frac{\log \varepsilon_d}{\varepsilon_d - \varepsilon_d^{-1}} e^{-a\left(\varepsilon_d^2 + \varepsilon_d^{-2}\right)}, \; a \to 0+, \tag{3.182}$$

where the sum is over the fundamental units of the theorem.

Exercise 3.7.33. Prove formulas (3.181) and (3.182).

Hint. *Replacing the sum over all hyperbolic conjugacy classes with that over primitive hyperbolic conjugacy classes is a standard argument in analytic number theory (see, e.g., Lang [388, p. 159]).*

One must finally make a Tauberian argument to complete the proof of Theorem 3.7.6. This time the argument is more complicated than that in the proof of Theorem 1.2.5 of Sect. 1.2. This argument in given in Exercise 3.7.34. □

Note that by Exercise 3.7.20, Theorem 3.7.6 can also be viewed as the prime geodesic theorem, giving the asymptotics of the length spectrum of closed geodesics on the Riemann surface $SL(2,\mathbb{Z})\backslash H$.

Exercise 3.7.34 (H. Stark). Complete the proof of Theorem 3.7.6 using the following hints.

Step 1. The ε_d^{-1} appearing in Eq. (3.182) can be thrown away since $\varepsilon_d \to \infty$ as $d \to \infty$.

Step 2. To apply Theorem 1.2.5 of Sect. 1.2, set

$$a(t) = \sum_{\varepsilon_d^2 \le t} h_d \frac{\log \varepsilon_d}{\varepsilon_d},$$

and obtain

$$a(t) \sim 2t^{1/2} \qquad \text{as } t \text{ approaches infinity.}$$

Step 3. Change variables via $t = v^2$ and obtain

$$\sum_{\varepsilon_d \le v} h_d \frac{\log \varepsilon_d}{\varepsilon_d} \sim 2v \qquad \text{as } v \text{ approaches infinity.}$$

Step 4. Given any $\delta > 0$, only the terms $x^{1-\delta} < \varepsilon_d < x$ matter and thus you obtain

$$A(x) = \sum_{\varepsilon_d \le x} \frac{h_d}{\varepsilon_d} \sim \frac{2x}{\log x}, \quad \text{as } x \text{ approaches } \infty.$$

Step 5. Use integration by parts to see

$$\sum_{\varepsilon_d \le x} h_d = \sum_{\varepsilon_d \le x} h_d \frac{\varepsilon_d}{\varepsilon_d} = \int_0^x t \, dA(t) \sim 2\frac{x}{\log x} x - 2\int_e^x \frac{t}{\log t} dt \sim \frac{x^2}{\log x^2}.$$

It is interesting to compare the result of Theorem 3.7.6 with a conjecture made by Gauss which was later proved by Siegel [600, Vol. II, pp. 473–491]:

$$\sum_{d \leq x} h_d \log \varepsilon_d \sim \frac{\pi^2}{18\zeta(3)} x^{3/2}, \quad x \to \infty. \tag{3.183}$$

In both formulas the ds run over positive integers which are not perfect squares and which are congruent to 0 or 1 modulo 4. It does not appear to be possible to use either Theorem 3.7.6 or formula (3.183) or both to attack the question of whether an infinite number of h_d are equal to 1.

To emphasize the analogies with prime number theory one can define **Selberg's zeta function** for $SL(2,\mathbb{Z})$ as

$$Z(s) = \prod_{k \geq 0} \prod_{d} (1 - \varepsilon_d^{-2s-2k})^{h_d} \quad \text{if } \operatorname{Re} s \geq 1. \tag{3.184}$$

This function has an analytic continuation as a meromorphic function in the whole complex s-plane with nontrivial zeros at s such that $s(s-1)$ is an eigenvalue in the discrete spectrum of Δ on $\Gamma\backslash H$ and at the poles of the function

$$\varphi(s) = \Lambda(1-s)/\Lambda(s), \; \Lambda(s) = \pi^{-s/2}\Gamma(s/2)\zeta(s)$$

in $\operatorname{Re} s > \frac{1}{2}$ (see Gangolli and Warner [195], Hejhal [261], Selberg [569], Venkov [693], Marie-France Vignéras [700] and Voros [710]). Vignéras [700] shows that the functional equation of Selberg's zeta function involves the Barnes double gamma function, which was used by Shintani [595] in the evaluation of zeta functions of totally real number fields at negative integer values.

Selberg's zeta function is shown by Elstrodt [155, p. 68] to arise through the formation of the trace of $R_\lambda R_\mu$, where $R_\lambda = (\Delta - \lambda)^{-1}$ = the resolvent of the Laplacian. Thus the Selberg zeta function actually contains as much information as the trace formula itself.

Ruelle [549] discusses various zeta functions from number theory and geometry. He notes that Selberg's zeta function is a special case of a zeta function considered by Smale for differentiable flows on compact differentiable manifolds, assuming that the fundamental domain $\Gamma\backslash H$ is compact, of course. Graph theory analogues of Selberg's zeta function, known as Ihara zeta functions, are discussed in my book [671].

There are many other applications of Selberg's trace formula. For example, one can return to the question discussed in the introduction to this section: Can one hear the shape of a drum? This is the first of the following questions. The second question can be viewed as involving the set of lengths of periodic billiard paths on the fundamental domain. The questions are theoretical rather than practical since it is unlikely that any human can hear all the overtones of a drum or compute all the periodic billiard lengths. Physicists have found the results to be of interest in work on quantum chaos. See Gutzwiller [242].

Questions About Vibrating Manifolds

1. Is the drum determined up to isometry by the spectrum of the Laplacian?
2. Is the set of lengths of closed geodesics determined by the spectrum of the Laplacian? You see the lengths and you hear the eigenvalues.
3. Is the drum determined (up to isometry) by its length spectrum?

Some Answers

1. Marie-France Vignéras [701] shows that there are non-isometric compact Riemann surfaces whose Laplacians have the same spectrum. Sunada [649] showed that one can obtain isospectral but nonisometric manifolds using the trace formula and a bit of group theory that goes back to some work in algebraic number theory giving examples of nonisomorphic algebraic number fields with the same Dedekind zeta function. We discuss graph theory analogues of this in Stark and Terras [633]. See Terras [668] for finite analogues, as well as [671]. Another reference for the compact case is Buser [74]. The Buser examples ultimately led to the planar drums that cannot be heard. See Carolyn Gordon et al. [227].
2. Gangolli [191] shows, using the Selberg trace formula, that when $\Gamma\backslash H$ is compact the spectrum of the Laplacian determines the lengths of closed geodesics. Huber [307] proved that the Laplace spectra of two compact Riemann surfaces are the same if and only if the surfaces have the same length spectra. See Buser [74, Chap. 9].
3. Wolpert [744,745] showed that a generic compact Riemann surface is determined up to isometry by its length or Laplace spectrum. See Buser [74, Chap. 10].

Selberg also used the trace formula to compute dimensions of spaces of holomorphic cusp forms (see Selberg [569] and Hejhal [261]). This method has been used with success for higher-rank groups (see Volume II or [667]). It is closely connected with Hirzebruch's generalization of the Riemann–Roch theorem as well as the Atiyah–Singer index theorem. However, there does not exist a formula for the dimensions of spaces of cuspidal Maass waveforms of the type we have considered.

Selberg [569] and Eichler [152] use the trace formula to obtain a formula involving class numbers of positive definite binary quadratic forms for the traces of Hecke operators (see Hejhal [261] and Lang [391], the chapter by Zagier on the trace formula, with correction in Zagier [753]). This version of the trace formula is called the **Eichler–Selberg trace formula**.

The trace formula is intrinsic to work on the Langlands program. Of course the language of this subject is adelic representation theory which we have avoided here. See Bernstein and Gelbart [37].

The trace formula appears in much of the work on the Artin conjecture on the holomorphy of Artin L-functions. References are Saito [552] and the article of Gérardin and Labesse in the Corvallis conference proceedings (see Borel and Casselman [52, Vol. II, pp. 115–134]), as well as the article of Shintani in the same conference proceedings [loc. cit., pp. 97–110], and finally the notes of Langlands

[394] in which a proof of a new case of Artin's conjecture is given. See Tunnell [686, 687] for an extension of Langlands' results.

We consider a final application of the trace formula in Sarnak's formula (3.185) below counting modular knots with given link number with respect to the trefoil knot.

3.7.6 Tables Summarizing the Main Results

Tables 3.11 and 3.12 plus Fig. 3.46 summarize the main results of this chapter—at least those which are easily summarized. We are assuming $\Gamma = SL(2,\mathbb{Z})$ throughout the tables. These should be compared with the tables in Sect. 3.7.8 where H is replaced by the finite upper half-plane H_q. Here we view $\mathbb{R}^2$ as the symmetric space $\mathbb{R}^2/\{0\}$ rather than the symmetric space of the Euclidean group $M\left(2,\mathbb{R}^2\right)$ modulo the subgroup $SO(2)$.

3.7.7 Modular Knots

There is a surprising connection between primitive closed geodesics in $SL(2,\mathbb{Z})\backslash H$ and knots coming from the Lorenz nonlinear differential equations introduced in 1963 as a model of atmospheric convection. These connections were found by Ghys [210]. See the beautiful website of Ghys and Leys [211]. For more information on knots see Adams [1].

First one can identify $SL(2,\mathbb{R})/SL(2,\mathbb{Z})$ with the complement of the trefoil knot in $\mathbb{R}^3$. The trefoil knot τ is the knot pictured in Fig. 3.47. To see this, one must identify $SL(2,\mathbb{R})/SL(2,\mathbb{Z})$ with the space of lattices $L = \mathbb{Z}\omega_1 + \mathbb{Z}\omega_1$ in the plane and look at the corresponding Eisenstein series

$$g_2(L) = \sum_{\omega \in L-\{0\}} \omega^{-4} \quad \text{and} \quad g_3(L) = \sum_{\omega \in L-\{0\}} \omega^{-6}.$$

Then (g_2, g_3) correspond to a lattice iff the discriminant $\Delta = g_2^3 - 27g_3^2 \neq 0$. So the space of lattices can be identified with $\mathbb{C}^2 - \{\Delta = 0\}$. Scale the lattice by a positive constant in order to restrict to lattices such that $|g_2|^2 + |g_3|^2 = 1$.

The flow in $SL(2,\mathbb{R})/SL(2,\mathbb{Z})$ given by $\begin{pmatrix} e^{t/2} & 0 \\ 0 & e^{-t/2} \end{pmatrix} \in SL(2,\mathbb{R})$ corresponds to the geodesic flow on $SL(2,\mathbb{Z})\backslash H$, where H denotes the upper half-plane. Primitive closed geodesics are orbits of this flow. They correspond to primitive hyperbolic conjugacy classes in $SL(2,\mathbb{Z})$; i.e., $\{\gamma\}$ with $|Tr(\gamma))| > 2$.

You can thus view the orbits of this flow in $SL(2,\mathbb{Z})\backslash H$ as knots in $\mathbb{R}^3 - \tau$ (the complement in 3-space of the trefoil knot). These are now known as **modular knots**. We use the notation k_γ for the modular knot associated to $\gamma \in SL(2,\mathbb{Z})$. For example,

Table 3.11 Comparison of Euclidean and non-Euclidean harmonic analysis in two dimensions

General	Euclidean	Non-Euclidean
Symmetric space $X \cong G/K$	$x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \in \mathbb{R}^2$ Euclidean plane	$z = x + iy \in H$ Poincaré upper half-plane
Arc length ds	$ds^2 = dx_1^2 + dx_2^2$	$ds^2 = y^{-2}\left(dx^2 + dy^2\right)$
Laplacian	$\Delta = \frac{\partial^2}{\partial x_1^2} + \frac{\partial^2}{\partial x_2^2}$	$\Delta = y^2 \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right)$
G-invariant area $d\mu$	$d\mu = dx_1 dx_2$	$d\mu = y^{-2}\, dx\, dy$
G=isometry group	$\mathbb{R}^2$	$g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL(2,\mathbb{R})$
K =subgroup of G fixing the origin	$K = \{0\}$	$SO(2)$, origin=i
Action of $g \in G$ on $x \in X$	$x \longmapsto x + g$, vector addition	$z \longmapsto gz = \frac{az+b}{cz+d}$
Γ =discrete subgroup of G	$\mathbb{Z}^2$	$SL(2,\mathbb{Z})$
Elementary eigenfunctions of Δ	$e_a(x) = e^{2\pi i\ {}^t ax}$	$p_s(z) = (\operatorname{Im} z)^s$
Eigenvalues	$\Delta e_a = -4\pi^2 \lVert a \rVert^2 e_a$	$\Delta p_s = s(s-1) p_s$
Spherical function (zonal)	$e_a(x)$	$h_s(z) = \int\limits_K (\operatorname{Im}(kz))^s dk$ $\Delta h_s = s(s-1) h_s$
Helgason–Fourier transform	$\hat{f}(y) = \int\limits_{\mathbb{R}^2} f(x)\overline{e_y(x)}dx$	$\mathcal{H}f(s,k) = \int\limits_H f(z)\overline{p_s(kz)}d\mu,$ $s \in \mathbb{C}, k \in K$
Fourier inversion or spectral decomposition of Δ	$f(x) = \hat{\hat{f}}(-x)$	$f(z) = \frac{1}{4\pi} \int\limits_{t \in \mathbb{R}} \int\limits_{k \in K} \mathcal{H}f(\frac{1}{2} + it, k)$ $\times p_{\frac{1}{2}+it}(kz) t \tanh(\pi t)\, dt\, dk$
Convolution defined by convolution on G	$\widehat{f * g} = \hat{f} \cdot \hat{g},\ f, g \in L^1(\mathbb{R}^2)$	$\mathcal{H}(f * g) = \mathcal{H}f \cdot \mathcal{H}g,$ $f, g \in L^1(H)$, for $K-$invariant g
Differentiation	$\widehat{\Delta f}(y) = -4\pi^2 \lVert y \rVert^2 \hat{f}(y)$	$\mathcal{H}(\Delta f)(s,k) = \overline{s}(\overline{s}-1)\mathcal{H}f(s,k)$
Heat equation $u_t = \Delta u,\ u(x,0) = f(x)$ $u(x,t) = f * g_t$ g_t= fundamental solution	$g_t(x) = \frac{1}{4\pi t} \exp\left(-\frac{\lVert x \rVert^2}{4t}\right)$	$g_t(ke^{-r}i)$ $= \frac{\sqrt{2}}{(4\pi t)^{3/2} e^{t/4}} \int\limits_r^\infty \frac{be^{-b^2/(4t)}}{\sqrt{\cosh b - \cosh r}} db$
Horocycle transform	No analogue	$F(y) = y^{-1/2} \int_{\mathbb{R}} f(x+iy) dx$ invertible

we have the green knot corresponding to the hyperbolic element $\begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}$ in Fig. 3.48, shown there winding about the trefoil knot. See Ghys and Leys [211] for this and more examples.

Ghys proved (see [210]) that such modular knots are Lorenz knots. The Lorenz knots are closed solution curves of the Lorenz equations:

$$\frac{dx}{dt} = 10(y - x), \frac{dy}{dt} = 28x - y - xz, \frac{dz}{dt} = xy - \frac{8}{3}z.$$

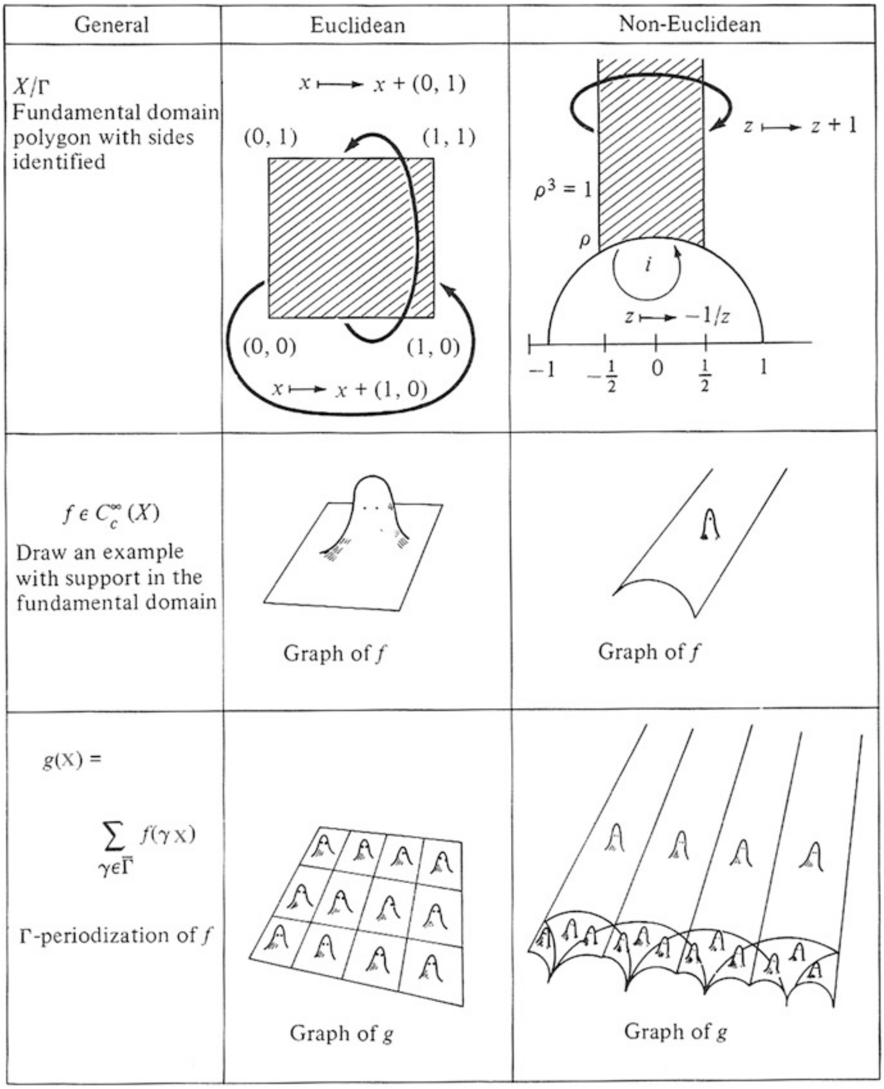
Table 3.12 Harmonic analysis on the fundamental domain

General	Euclidean	Non-Euclidean
Complete orthonormal set of eigenfunctions of Δ on X/Γ	$e_a(x) = \exp(2\pi i \, {}^t a x)$, $a \in \mathbb{Z}^2$ purely discrete spectrum	**Continuous spectrum** of Eisenstein series $E_s(z) = \sum_{\Gamma_\infty \backslash \Gamma} p_s(\gamma z), \ \ \mathrm{Re}\, s > 1,$ with analytic continuation to other values of s **Discrete spectrum** of cusp forms and constants $v_0 = \sqrt{\frac{3}{\pi}}, \quad \{v_n\}_{n \geq 1},$ complete orthonormal set
Eigenvalues	$\Delta e_a = -4\pi^2 \|a\|^2 e_a$	$\Delta E_s = s(s-1)E_s, \ \ \Delta v_0 = 0,$ $\Delta v_n = s_n(s_n - 1) v_n$
Spectral resolution of Δ on $L^2(X/\Gamma)$	$g(x) = \sum_{a \in \mathbb{Z}^2} (g, e_a) e_a(x)$ ordinary Fourier series	$g(z) = \sum_{n \geq 0} (g, v_n) v_n(z)$ $+ \frac{1}{4\pi i} \int_{\mathrm{Re}\, s = \frac{1}{2}} (g, E_s) E_s(z) ds$ Roelcke–Selberg spectral decomposition of Δ on $L^2(SL(2,\mathbb{Z}) \backslash H)$
Poisson summation formula $f \in C_c^\infty(K \backslash X)$ $g(x) = \sum_{\gamma \in \Gamma} f(\gamma x)$ $= \sum_{\alpha \in A} \int \hat{f}(\alpha) \overline{e_\alpha(o)} e_\alpha(x) \sigma(\alpha) d\alpha$ This is also Mercer's theorem.	$\sum_{a \in \mathbb{Z}^2} f(x+a) = \sum_{a \in \mathbb{Z}^2} \hat{f}(a) e_a(x)$ Here $e_a(0) = 1$. $\hat{f}(a) = \int_{\mathbb{R}^2} f(x) \overline{e_a(x)} dx$	$\sum_{\gamma \in \Gamma/\{\pm I\}} f(\gamma x) = \sum_{n \geq 0} \hat{f}(s_n) \overline{v_n(i)} v_n(z)$ $+ \frac{1}{4\pi i} \int_{\mathrm{Re}\, s = \frac{1}{2}} \hat{f}(s) \overline{E_s(i)} E_s(z) ds$ $\hat{f}(s) = \int_H f(z) \overline{p_s(z)} d\mu$

(continued)

Table 3.12 (continued)

General	Euclidean	Non-Euclidean
Conjugacy classes $\subset \Gamma$ $\{\gamma\} = \{x\gamma x^{-1} \vert x \in \Gamma\}$ Centralizer $\Gamma_\gamma = \{x \in \Gamma \vert x\gamma = \gamma x\}$	Abelian group $\vert\{g\}\vert = 1$ everything central	Central $\{\pm I\}$; Hyperbolic $\sim \pm \begin{pmatrix} t & 0 \\ 0 & t^{-1} \end{pmatrix}, t > 1$ $\Gamma_\gamma = \langle \gamma_0 \rangle$ cyclic, γ_0 primitive hyperbolic; Plus parabolic & elliptic
Selberg trace formula	Poisson summation	$\sum_{\lambda_n = s_n(1-s_n)} \hat{f}(s_n) = \text{area}(\Gamma\backslash H) f(i)$ $+ \sum_{\{\gamma\}\text{ hyperbolic}} \frac{\log N\gamma_0}{N\gamma^{\frac{1}{2}} - N\gamma^{-\frac{1}{2}}} F(N\gamma)$ $+$parabolic & elliptic terms
Selberg zeta function	No analogue	$Z(s) = \prod_{\substack{\{\gamma_0\} \\ \text{primitive} \\ \text{hyperbolic}}} \prod_{j \geq 0} (1 - N\gamma_0^{-s-j})$ nontrivial zeros correspond to spectrum Δ on $L^2(\Gamma\backslash G/K)$ except for finite # zeros, RH true
Application to circle problem	$N_{\mathbb{Z}^2}(x)$ $= \#\left\{ n \in \mathbb{Z}^2 \;\middle\vert\; d(n,0)^2 \leq x \right\}$ $\sim \pi x, \text{ as } x \to \infty$ $d(n,0)^2 = \Vert n \Vert^2 = n_1^2 + n_2^2$	$N_\Gamma(x) =$ $\#\{ \gamma \in \Gamma \mid \cosh d(\gamma i, i) \leq x\}$ $\sim 6x, \text{ as } x \to \infty$ $d(z,w) =$ non-Euclidean distance z to w $2\cosh d(\gamma i, i) = Tr({}^t\gamma\gamma)$
Theta function	$\theta_{\mathbb{Z}^2}(a) = \sum_{n \in \mathbb{Z}^2} \exp(-a^t nn)$ $= \frac{\pi}{a} \theta_{\mathbb{Z}^2}(1/a) \sim \frac{\pi}{a}, \text{ as } a \to 0+$	$\theta_\Gamma(a) = \sum_{\gamma \in \Gamma/\{\pm I\}} \exp\left(\frac{-a}{2} Tr({}^t\gamma\gamma)\right)$ $f_a(z) = \exp(-a \cosh d(i,z))$ $\widehat{f_a}(s) = 2\sqrt{\frac{2\pi}{a}} K_{\bar{s}-\frac{1}{2}}(a)$ $\theta_\Gamma(a) \sim \frac{6}{a}, \text{ as } a \to 0+$



General	Euclidean	Non-Euclidean
X/Γ Fundamental domain polygon with sides identified	$x \mapsto x + (0, 1)$ (0, 1) (1, 1) (0, 0) (1, 0) $x \mapsto x + (1, 0)$	$z \mapsto z + 1$ $\rho^3 = 1$ ρ i $z \mapsto -1/z$ -1 $-\frac{1}{2}$ 0 $\frac{1}{2}$ 1
$f \in C_c^\infty(X)$ Draw an example with support in the fundamental domain	Graph of f	Graph of f
$g(X) = \sum_{\gamma \in \overline{\Gamma}} f(\gamma X)$ Γ-periodization of f	Graph of g	Graph of g

Fig. 3.46 Γ-periodizations of compactly supported functions on H

Integrate these ODEs and obtain orbits that give a flow $u(t), t \in \mathbb{R}$. These solution orbits stay in the neighborhood of a fractal region shaped like a butterfly (called a strange attractor). A Lorenz knot is a periodic orbit of the Lorenz flow. These knots were studied by Joan Birman and Robert Williams in 1983 (see [41]). In particular, they found the link types of these knots. It is hard to compute such things thanks to the chaotic nature of the Lorenz equations and the butterfly effect, which means that a slight change in initial condition can make a huge change in the solution.

To define the linking number of two oriented nonintersecting knots, we need Fig. 3.49. Project the knots onto the plane. At the points where the purple knot

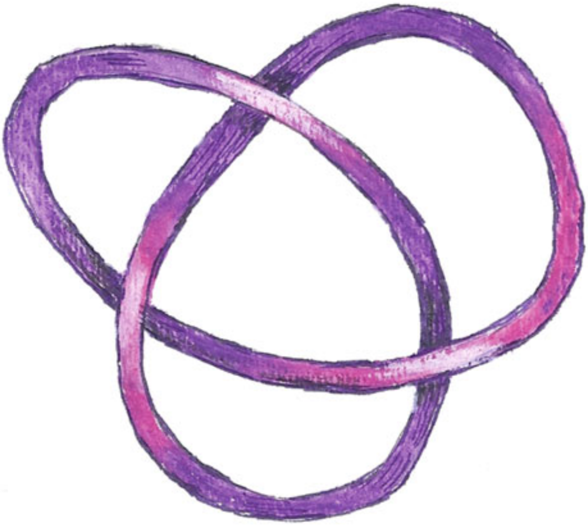


Fig. 3.47 The trefoil knot. τ

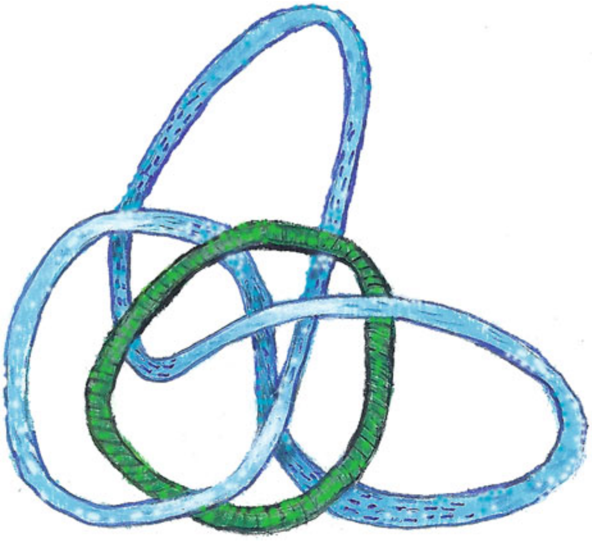
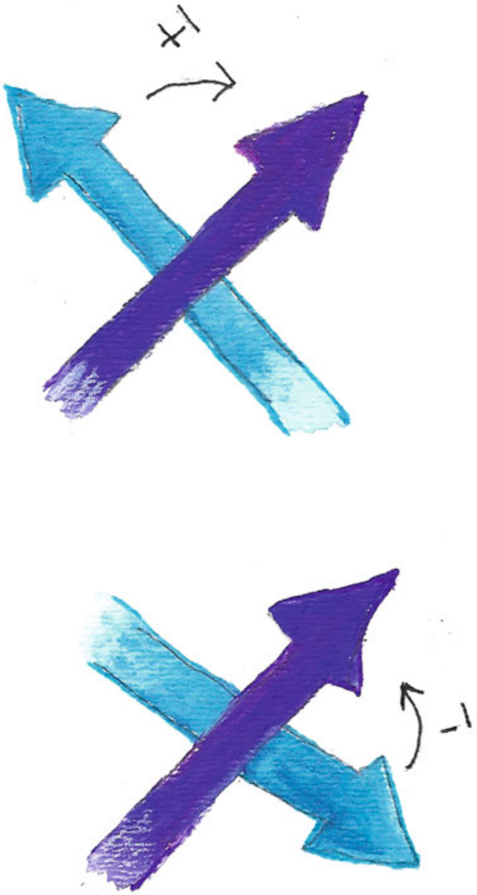


Fig. 3.48 The modular knot associated with $\gamma = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}$ (*pictured in green*) from Ghys and Leys [211] shown linked with the trefoil knot (*blue*)

Fig. 3.49 Computing the linking number of the purple knot with the turquoise knot



crosses over the turquoise one, you get two types of picture. If you see the top picture in Fig. 3.49, you get $+1$. In this case, when you rotate the knot that goes under; i.e., the turquoise knot, **clockwise** so that it lines up with the purple knot, then the arrows line up. If you see the bottom picture in Fig. 3.49, you get -1. In this case, you need to rotate the turquoise knot in a **counterclockwise** direction to get the arrows to line up. Sum these numbers over all crossings of the purple knot over the turquoise knot and get the **linking number** denoted $\mathrm{link}(\mathrm{purpleknot}, \mathrm{turquoiseknot})$. Adams [1] shows on p. 20 that it does not matter what projection you choose to compute the linking number. He also shows that this number can be 0 even though the knots cannot be pulled apart.

Now comes the amazing part. Since the modular knots can be viewed as occurring in the complement of the trefoil knot in the unit ball in $\mathbb{R}^3$, it is natural to want to compute the linking numbers of these modular knots with the trefoil knot using modular forms somehow. Ghys proved that the linking number of a modular knot with the trefoil knot is the Rademacher function $\Psi : PSL(2,\mathbb{Z}) = SL(2,\mathbb{Z})/\{\pm I\} \to \mathbb{Z}$, which comes out of the multiplier system for the Dedekind eta function; i.e.,

$$\mathrm{link}(k_\gamma, \tau) = \Psi(\gamma).$$

Here the Rademacher function $\Psi(\gamma)$ comes from the transformation formula for the eta function:

$$\begin{aligned} \eta(\gamma z) &= v_{\frac{1}{2}}(\gamma)(cz+d)^{\frac{1}{2}}\eta(z), \text{ for all } \quad \gamma \in SL(2,\mathbb{Z}), \\ v_{\frac{1}{2}}(\gamma) &= e^{i\pi\Psi(\gamma)/12}, \text{ when } Tr(\gamma) > 0. \end{aligned}$$

The computation of the Rademacher function can be done by writing $\gamma \in SL(2,\mathbb{Z})$ as a product of $U = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ and $V = \begin{pmatrix} 1 & -1 \\ 1 & 0 \end{pmatrix}$. If $\gamma = UV^{e_1} \cdots UV^{e_r}$, then $\Psi(\gamma) = e_1 + \cdots + e_r$. See Rademacher and Grosswald [527, pp. 45–58] for proofs of this and much more information on the Rademacher function and its connection to Dedekind sums.

Exercise. Compute $\mathrm{link}(k_\gamma, \tau)$ for $\gamma = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}$.

Sarnak [559] uses the Selberg trace formula to show that

$$\#\left\{ \{\gamma\} \,\middle|\, \begin{array}{l} \text{primitive hyperbolic in } SL(2,\mathbb{Z}) \\ Tr(\gamma) \le x\,, \text{ and } \mathrm{link}(k_\gamma, \tau) = n \end{array} \right\} \sim \frac{x^2}{12\,(\log x)}, \quad \text{as} \quad x \to \infty. \tag{3.185}$$

He also shows a more precise version of this formula which implies that the most common linking number between a modular knot and the trefoil knot is 0.

One wonders whether other topological invariants of modular knots and links can be computed using modular forms somehow.

3.7.8 Finite Analogue of the Selberg Trace Formula

It is hard to resist seeking a trace formula for some subgroup Γ of $GL(2,\mathbb{F}_q)$ acting on the finite upper half-plane H_q, if $q = p^r$, for an odd prime p and $r > 1$. We did this in [668] for $\Gamma = GL(2,\mathbb{F}_p)$. Shaheen [583, 584] considered the trace formula for various other subgroups Γ of G. Here we shall only state the trace formula for $\Gamma = GL(2,\mathbb{F}_p)$ acting on H_{p^2}.

In our first table we summarize what we know from our earlier section on finite upper half-planes and finite upper half-plane graphs. We found in our book [668] that it was necessary to know the irreducible unitary representations of $G = GL(2,\mathbb{F}_q)$, $q = p^r$, p an odd prime, since not only principal series but also discrete series representations arise in studying the eigenvalues of the adjacency operators

Table 3.13 Finite upper half-planes

Finite upper half-plane
Space $= H_q$, q odd $= \{z = x+\sqrt{\delta}y \mid x,y \in \mathbb{F}_q, y \neq 0\}$ where $\delta \neq a^2$, for any $a \in \mathbb{F}_q$
Group $G = GL(2,\mathbb{F}_q)$ $= \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \middle\vert\, ad-bc \neq 0 \right\}$
Group action $gz = \frac{az+b}{cz+d}$, $z \in H_q$
Origin = $\sqrt{\delta}$
Subgroup K fixing origin $= K = \left\{ \begin{pmatrix} a & b\delta \\ b & a \end{pmatrix} \in G \right\} \cong \mathbb{F}_q(\sqrt{\delta})^*$
$H_q \cong G/K$
Pseudo-distance $d(z,w) = \frac{N(z-w)}{\operatorname{Im} z \operatorname{Im} w}$, $\operatorname{Im}(x+y\sqrt{\delta}) = y$, $N(x+y\sqrt{\delta}) = x^2-y^2\delta$
Graph $X_a = X_q(\delta,a)$, vertices $z \in H_q$ edge between z and w if $d(z,w) = a$
A_a = **adjacency operator** for X_a
$\Delta_a = A_a - (q+1)I$, if $a \neq 0$ or 4δ
Spherical function $h_\pi(z) = \frac{1}{\lvert K \rvert} \sum_K \chi_\pi(kz)$, $\pi \in \hat{G}^K$, $\chi_\pi = Tr(\pi)$ *i.e.*, π occurs in $Ind_K^G 1$, $d_\pi = \deg \pi$
Spherical transform of $f : K\backslash G/K \longrightarrow \mathbb{C}$ $\hat{f}(\pi) = \sum_{z \in H_q} f(z)\overline{h_\pi(z)}$
Inversion $f(z) = \frac{1}{\lvert G \rvert} \sum_{\pi \in \hat{G}^K} d_\pi \hat{f}(\pi) h_\pi(z)$
Horocycle transform $F(y) = \sum_{x \in \mathbb{F}_q} f(x+y\sqrt{\delta})$ not invertible

on finite upper half-plane graphs. We need the discrete series representations again in our study of the trace formula. The notation $\rho = Ind_K^G 1$ means the induced representation obtained by inducing the trivial representation on K up to the representation ρ on G by setting $(\rho(g)f)(x) = f(gx)$, for all $x, g \in G$. Here $f \in V = \{f : G \to \mathbb{C} \mid f(kx) = f(x), \forall k \in K, x \in G\}$. The notation $\hat{G}^K$ means the irreducible unitary representations of G that occur in $Ind_K^G 1$.

The details of the computations for the following two tables are to be found in [668]. Table 3.13 summarizes the basic facts about finite upper half-planes and their associated graphs. Table 3.14 summarizes the trace formula for the subgroup $\Gamma = GL(2,\mathbb{F}_p)$, p an odd prime acting on H_{p^2}. These tables should be compared with the tables in Sect. 3.7.6.

It is also possible to consider p-adic symmetric spaces, which are $p+1$-regular trees; i.e., graphs that have no cycles. This is done in my book [668, Chap. 24] plus

Table 3.14 Selberg trace formula on finite upper half-plane

$\Gamma \subset G = GL(2,\mathbb{F}_q);\quad \Gamma = GL(2,\mathbb{F}_p),\quad q = p^2$	
Conjugacy classes in Γ	
Central	$\begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix}, a \in \mathbb{F}_p^*$
Hyperbolic	$\begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix}, a \neq b \in \mathbb{F}_p^*$
Parabolic	$\begin{pmatrix} a & 1 \\ 0 & a \end{pmatrix}, a \in \mathbb{F}_p^*$
Elliptic	$\begin{pmatrix} a & b\xi \\ b & a \end{pmatrix},\ \begin{matrix} a,b \in \mathbb{F}_p, b \neq 0, \\ \xi \neq u^2, u \in \mathbb{F}_p \end{matrix}$
Selberg trace formula $(q = p^2)$	
$\sum_{\pi \in \hat{G}^K} \hat{f}(\pi) mult(\pi, Ind_\Gamma^G 1)$ $= \lvert\Gamma\backslash G\rvert(p-1)f(\sqrt{\delta}) + \frac{(q+1)(q-1)^2}{2(p-1)} \sum_{\substack{c \in \mathbb{F}_p^* \\ c \neq 1}} F(c)$ $+ \frac{q(q^2-1)}{p} \left\{F(1) - f(\sqrt{\delta})\right\} + \frac{q^2-1}{2} \sum_{\substack{a,b \in \mathbb{F}_p \\ b \neq 0}} F\left(\frac{a+b\eta}{a-b\eta}\right),$ where $\eta^2 = \xi,\ \eta \in \mathbb{F}_q$	

the papers with Dorothy Wallace [672] and with Horton and Newland [304]. Tables comparing the tree with the upper half-plane and the finite upper half-plane are to be found in my paper [669]. The trace formula on a tree has many applications; e.g., to obtain the Ihara determinant formula for the Ihara zeta function of a finite regular graph. This is discussed in [668, Chap. 24] and [304].

Another version of the trace formula on finite upper half-planes was found by Shaheen in [583, 584]. Shaheen follows the methods of Hejhal [261, Vol. I], avoiding all mention of group representations.

References

1. C. C. Adams, *The Knot Book*, Freeman, N.Y., 1994.
2. L. M. Adleman, C. Pomerance, and R.S. Rumely, On distinguishing prime numbers from composite numbers, *Annals of Math, (2), 117* (1983), 173–206.
3. L. Ahlfors, *Complex Analysis*, McGraw-Hill, N.Y., 1979.
4. M. Alsina and P. Bayer, *Quaternion orders, quadratic forms, and Shimura curves, volume 22 of CRM Monograph Series*, American Mathematical Society, Providence, RI, 2004.
5. J. Altman, *Microwave Circuits*, Van Nostrand, Princeton, NJ, 1964.
6. R. J. Anderson, On the Mertens conjecture for cusp forms, *Mathematika, 26* (1979), 236–249; *27* (1980), 261.
7. T. W. Anderson, *An Introduction to Multivariate Statistical Analysis,* Wiley, NY, 1958.
8. A. N. Andrianov, On zeta functions of Rankin type associated with Siegel modular forms, *Lecture Notes in Math., 627*, Springer-Verlag, NY, 1977, 325–338.
9. J. Angel, Finite upper half planes over finite fields, *Finite Fields Applics., 2* (1996), 62–86.
10. T. Apostol, *Calculus, Vols. I and II*, Blaisdell, Waltham, MA, 1967.
11. T. Apostol, *Mathematical Analysis*, Addison-Wesley, NY, 1974.
12. T. Apostol, *Modular Functions and Dirichlet Series in Number Theory*, Springer-Verlag, NY, 1976.
13. G. Arfken, *Mathematical Methods for Physicists*, Academic, NY, 1970.
14. E. Artin, *Collected Papers*, Addison-Wesley, Reading, MA, 1965.
15. J. M. Ash (Ed.), *Studies in Harmonic Analysis, Vol. 13, M.A.A. Studies in Math.*, Math. Assoc. of America, Wash., D.C., 1976.
16. A. O. Atkin and J. Lehner, Hecke operators on $\Gamma_0(N)$, *Math. Ann., 185* (1970), 134–160.
17. L. Auslander, *Differential Geometry*, Harper and Row, N.Y., 1967.
18. L. Auslander et al., *Flows on Homogeneous Spaces*, Princeton U. Press, Princeton, NJ, 1963.
19. G. Backus and F. Gilbert, The rotational splitting of the free oscillations of the Earth, *Proc. Natl. Acad. Sci., 47* (1961), 362–371.
20. A. J. Baden Fuller, *Microwaves*, Pergamon, Oxford, UK, 1979.
21. A. Baker, *Transcendental Number Theory,* Cambridge U. Press, Cambridge, 1975.
22. E. Bannai, Character tables of commutative association schemes, in *Finite Geometries, Buildings, and Related Topics*, (W. M. Kantor, et al, Eds.), Clarendon Press, Oxford, 1990, pp. 105–128.
23. D. Barbasch, Fourier inversion for unipotent invariant integrals, *Trans. Amer. Math. Soc., 249* (1979), 51–83.
24. D. Barbasch, Fourier transforms of some invariant distributions on a semi-simple Lie group, *Lecture Notes in Math., 728*, Springer-Verlag, NY, 1979, 1–7.

A. Terras, *Harmonic Analysis on Symmetric Spaces—Euclidean Space, the Sphere, and the Poincaré Upper Half-Plane*, DOI 10.1007/978-1-4614-7972-7,
© Springer Science+Business Media New York 2013

25. V. Bargmann, Irreducible unitary representations of the Lorentz group, *Ann. of Math., 48* (1957), 568–640.
26. T. Barnet-Lamb, D. Geraghty, M. Harris, and R. Taylor, A family of Calabi-Yau varieties and potential automorphy II, *Publ. Res. Inst. Math. Sci., 47* (2011), no. 1, 29–98.1663–4926.
27. P. Barrucand, H. Williams, and L. Baniuk, A computational technique for determining the class number of a pure cubic field, *Math. Comp., 30* (1976), 312–323.
28. A. O. Barut and R. Rączka, *Theory of Group Representations and Applications*, Polish Scientific Pub., Warsaw, Poland, 1977.
29. H. Bass, The congruence subgroup problem, in TA. Springer (Ed.), *Proc. of Conf. in Local Fields*, Springer-Verlag, NY, 1967, 16–22.
30. P. T. Bateman and E. Grosswald, On Epstein's zeta function, *Acta Arith., 9* (1964), 365–373.
31. A. F. Beardon, *The Geometry of Discrete Groups*, Springer, NY, 1983.
32. T. Bedford, M. Keane and C. Series, *Ergodic Theory, Symbolic Dynamics and Hyperbolic Spaces*, Oxford U. Press, Oxford, UK, 1991.
33. R. Bellman, *A Brief Introduction to Theta Functions*, Holt, Rinehart and Winston, NY, 1961.
34. J. J. Benedetto, *Harmonic Analysis and its Applications,* CRC Press, Boca Raton, FL, 1997.
35. J. J. Benedetto and M. W. Frazier, *Wavelets: Mathematics and Applications*, CRC Press, Boca Raton, FL, 1994.
36. M. Berger, Geometry of the spectrum, I, *Proc. Symp. Pure Math., Vol. 27, Pt. II.,* Amer. Math. Soc., Providence, RI, 1975, 129–152.
37. J. Bernstein and S. Gelbart, *An Introduction to the Langlands Program*, Birkhäuser, Boston, MA, 2004.
38. A. T. Bharucha-Reid, *Probabilistic Methods in Applied Math., Vol. 1,* Academic, NY, 1968.
39. L. C. Biedenharn and J. D. Louck, *Angular Momentum in Quantum Physics, Theory and Application,* Addison-Wesley, Reading, MA, 1981.
40. J. L. Birman, Theory of crystal space groups and infra-red Raman lattice processes of insulating crystals, in S. Flugge, (Ed.), *Encyl. of Physics, Vol. XXV/2b*, Springer-Verlag, NY, 1974.
41. J. Birman and R. Williams, Knotted periodic orbits in dynamical systems. I. Lorenz's equations, *Topology* 22 (1983), 47–82.
42. P. Bloomfield, *Fourier Analysis of Time Series,* Wiley, NY, 1976.
43. S. Bochner, *Lectures on Fourier Integrals*, Princeton U. Press, Princeton, NJ, 1959.
44. S. Bochner, Review of L. Schwartz, Théorie des distributions, *Bull. Amer. Math. Soc., 58* (1952), 78–85.
45. H. Boerner, *Group Representations*, Springer-Verlag, NY, 1955.
46. E. Bogomolny, Quantum and arithmetical chaos, *ArXiv*, 23 Dec. 2003.
47. J. Bolte and F. Steiner, Eds., *Hyperbolic Geometry and Applications in Quantum Chaos and Cosmology,* Cambridge U. Press, Cambride, UK, 2012.
48. A. R. Booker, A, Strömbergsson, and A. Venkatesh, Effective Computation of Maass Cusp Forms, InternationalMathematics Research Notices, Volume 2006, Article ID 71281, Pages 1–34
49. R. E. Borcherds, "What is Moonshine" , *Proc. Internat. Congress Math. (Berlin, 1998),* pp. 607–615 (Doc. Math. Extra Vol. 1.
50. J. Border, *Nonlinear Hardy Spaces and Electrical Power Transfer, Ph.D. Thesis*, U.C.S.D., 1979.
51. A. Borel, Arithmetic properties of algebraic groups, *Proc. Internatl. Cong. Math., Stockholm, 1962*.
52. A. Borel and W. Casselman, *Automorphic Forms, Representations, and L-Functions, Proc. Symp. Pure Math., 33, A.M.S.*, Providence, RI, 1979.
53. A. Borel, S. Chowla, et al., *Seminar on Complex Multiplication, Lecture Notes in Math, 21*, Springer-Verlag, N.Y., 1966.
54. A. Borel and G. Mostow, *Algebraic Groups and Discontinuous Subgroups, Proc. Symp. Pure Math., 9,* A.M.S., Providence, RI, 1966.

55. A. Borel and N. Wallach, *Continuous Cohomology, Discrete Subgroups, and Representations of Reductive Groups, Math. Surveys and Monographs, Vol. 67*, Amer. Math. Soc., Providence, RI, 2000.
56. Z. I. Borevitch and I. R Shafarevitch, *Number Theory*, Academic, NY, 1966.
57. M. Born and K. Huang, *Dynamical Theory of Crystal Lattices,* Oxford U. Press, London, 1956.
58. P. Bougerol, Comportement asymptotique des puissances de convolution d'une probabilité sur un espace symétrique, *Astérisque, 74* (1980), 29–45.
59. P. Bousquet, *Spectroscopy and its Instrumentation,* Hilger, London, 1971.
60. R. Bowen, The equidistribution of closed geodesics, *Amer. J. Math., 94* (1972), 413–423.
61. R. Bracewell, *The Fourier Transform and its Applications*, McGraw-Hill, NY, 1965.
62. R. Bracewell, *Fourier Analysis and Imaging*, Kluwer, NY, 2003.
63. O. Bratteli and P. E. T. Jorgensen, *Wavelets Through a Looking Glass: TheWorld of the Spectrum*, Birkhauser Boston, Boston, MA, 2002.
64. M. Braun, *Differential Equations and Their Applications*, 4th Ed., Springer, NY, 1993.
65. R. P. Brent, On the zeros of the Riemann zeta function, *Math. Comp., 33* (1979), 1361–1372.
66. C. Breuil, B. Conrad, F. Diamond and R.Taylor, On the modularity of elliptic curves over Q: wild 3-adic exercises, *J. Amer. Math. Soc., 14* (2001), 843–939.
67. E. O. Brigham, *The Fast Fourier Transform,* Prentice-Hall, Englewood Cliffs, N.J., 1974.
68. R. W. Bruggeman, *Fourier Coefficients of Automorphic Forms, Lecture Notes in Math., 865*, Springer-Verlag, NY, 1981.
69. R. W. Bruggeman, Fourier coefficients of cusp forms, *Inv. Math., 45* (1978), 1–18.
70. R. W. Bruggeman, *Kuznetsov's Proof of the Ramanujan-Petersson conjecture for modular Forms of Weight Zero*, Math. Inst., Rijksuniversiteit, Utrecht (8–1979). (An error has since been found in the proof.)
71. D. Bump, *Automorphic Forms and Representations*, Cambridge U. Press, Cambridge, UK, 1997.
72. H. Burkhardt, Trigonometrische Reihen und Integrale (bis etwa 1850), *Enzycl. der Math. Wissen.,* Teubner, Leipzig, 1916.
73. R. Burridge and G. Papanicolaou, The geometry of coupled mode propagation in one- dimensional random media, *Comm. Pure Appl. Math., 25* (1972), 715–757.
74. P. Buser, *Geometry and Spectra of Compact Riemann Surfaces*, Birkhäuser, Boston, 1992.
75. L. Carleson, On the convergence and growth of partial sums of Fourier series, *Acta Math., 116* (1966), 135–157.
76. G. F. Carrier, M. Krook, and C. E. Pearson, *Functions of a Complex Variable*, McGraw-Hill, NY, 1966.
77. P. Cartier, Some numerical computations relating to automorphic functions, in A.O. Atkin and B. Birch (eds.), *Computers in Number Theory*, Academic, NY, 1971.
78. P. Cartier and D. Hejhal, Sur les zéros de la fonction zêta de Selberg, Inst. des Hautes Études Sci., preprint (1979).
79. J. W. S. Cassels and A. Fröhlich, *Algebraic Number Theory,* Thompson, Washington, D.C., 1967.
80. G. W. Castellan, *Physical Chemistry*, Addison-Wesley, Reading, MA, 1971.
81. T. Ceccherini-Silberstein, F. Scarabotti, and F. Tolli, *Harmonic Analysis on Finite Groups*, Cambridge U. Press, Cambridge, 2008.
82. N. Celniker, S. Poulos, A. Terras, C. Trimble, and E. Velasquez, Is there life on finite upper half planes?, *Contemp. Math.*, Vol. 143 (1993), 65–88.
83. S. Chandrasekhar, *Hydrodynamic and Hydromagnetic Stability,* Oxford U. Press, London, UK, 1961.
84. I. Chavel, *Eigenvalues in Riemannian Geometry*, Academic Press, Orlando, 1984.
85. C. Chevalley, *Theory of Lie Groups,* Princeton U. Press, Princeton, NJ, 1946.
86. P. Chiu, Covering with Hecke points, *Journal of Number Theory, 53* (1995) 25–44.

87. Y. Choquet-Bruhat, C. Dewitt-Morette, and M. Dillard-Bleick, *Analysis, Manifolds, and Physics*, North-Holland, NY, 1977.
88. S. Chowla and A. Selberg, On Epstein's zeta function, *J. Reine und Angew. Math., 227* (1978), 86–110.
89. R. V. Churchill and J. W. Brown, *Fourier Series and Boundary Value Problems*, McGraw- Hill, NY, 1978.
90. B. Cipra, Parlez-vous Wavelets? in *What's Happening in the Mathematical Sciences, Vol. 2* (1994), AMS, Providence, RI, 1994, pp. 23–26.
91. P. Clark and J. Voight, Algebraic curves uniformized by congruence subgroups of triangle groups, preprint.
92. J. L. Clerc and B. Roynette, Un théorème central limite, *Lecture Notes in Math., 739*, Springer-Verlag, NY, 1979, 122–131.
93. L. Clozel, M. Harris, R. Taylor, Automorphy for some l-adic lifts of automorphic mod l Galois representations. With Appendix A, summarizing work of R. Mann, and Appendix B by M.-F. Vignéras, *Publ.Math. Inst.Hautes Études Sci., No. 108* (2008), 1–181.
94. J. Coates, p-adic L-functions and Iwasawa's theory, in *Algebraic Number Fields,* A. Frohlich (Ed.), Academic, London, 1977, 269–354.
95. H. Cohen, *A Course in Computational Algebraic Number Theory*, Springer, Berlin, 1993.
96. J. E. Cohen, Ergodic theorems in demography, *Bull. Amer. Math. Soc. 1*, (1979), 275– 295.
97. H. Cohn, *A Classical Invitation to Algebraic Numbers and Class Fields*, Springer-Verlag, NY, 1978.
98. R. R. Coifman and G. Weiss, Representations of compact groups and spherical harmonics, *L'enseignement Math., 14* (1968), 123–173.
99. R. E. Collin, *Foundations for Microwave Engineering*, McGraw-Hill, N.Y., 1966.
100. E. U. Condon and H. Odabaşi, *Atomic Structure*, Cambridge U. Press, Cambridge, UK, 1980.
101. E. U. Condon and G. H. Shortley, *Theory of Atomic Spectra*, Cambridge U. Press, Cambridge, UK, 1935.
102. A. Connes, Trace formula in noncommutative geometry and the zeros of the Riemann zeta function, *Selecta Math. (N. S.) 5* (1999), 29–106.
103. A. Connes, Trace formula on the adèle class space and Weil positivity, in Curent developments in mathematics, 1997, *Intl. Press*, Boston, MA, 1999, pp. 5–64.
104. A. Connes, Noncommutative geometry and the Riemann zeta function, in *Mathematics: Frontiers and Perspectives*, Amer. Math. Soc., Providence, RI, 2000, pp. 35–54.
105. J. H. Conway, Monsters and moonshine, *Math. Intelligencer, 2* (1980), 165–171.
106. J. H. Conway and S.P. Norton, Monstrous moonshine, *Bull. London Math. Soc., 11* (1979), 308–339.
107. J. H. Conway, A. M. Odlyzko, and N. J. A. Sloane, Extreme self-dual lattices exist only in dimensions 1 to 8, 12, 14, 15, 23, and 24, *Mathematika, 25*, (1978), 36–43.
108. J. W. Cooley and J. W. Tukey, An algorithm for the machine calculation of complex Fourier series, *Math. Computation, 19* (1965), 297–301.
109. E. T. Copson, *Theory of Functions of a Complex Variable*, Oxford U. Press, Oxford, UK, 1960.
110. COSRIMS, *The Mathematical Sciences: A Collections of Essays*, M.I.T. Press, Cambridge, MA., 1969.
111. R. Courant and D. Hilbert, *Methods of Mathematical Physics, Vol. I,* Wiley-Interscience, N.Y., 1961.
112. J.-M. Couveignes and B. Edixhoven, Eds., *Computational Aspects of Modular Forms and Galois Representations*, Princeton U. Press, Princeton, NJ, 2011.
113. H. S. M. Coxeter, *Non-Euclidean Geometry*, U. of Toronto Press, Toronto, Canada, 1965.
114. H. S. M. Coxeter, *Introduction to Geometry,* Wiley, NY, 1961.
115. H. Cramér, *Mathematical Methods of Statistics,* Princeton U. Press, Princeton, N.J., 1946.
116. Yu. L. Daletskii, Continual integrals related to operator evolution equations, *Russ. Math. Surveys, 17* (1962), 3–115.

117. I. Daubechies, Orthonormal bases of compactly supported wavelets, *Comm. Pure Appl. Math., 41* (1988), 909–996.
118. I. Daubechies, *Ten Lectures on Wavelets, CBMS-NSF Regional Conf. Ser. in Appl. Math.*, SIAM, Philadelphia, PA, 1992.
119. I. Daubechies and J. C. Lagarias, Two-scale difference equations I. Global regularity of solutions, *SIAM J. Math. Anal.,* 22 (1991), pp. 1388–1410.
120. H. Davenport, *Multiplicative Number Theory*, Springer-Verlag, NY, 1981.
121. H. Davenport and H. Heilbronn, On the zeros of certain Dirichlet series I, II, *J. London Math. Soc., 11* (1936), 181–185, 307–312.
122. S. P. Davis and M. C. Abrams, and J. W. Brault, *Fourier Transform Spectrometry*, Academic Press, San Diego,CA, 2001.
123. N. G. de Bruijn, Algebraic theory of Penrose's nonperiodic tilings of the plane, I, II, *Nederl. Akad. Wetensch. Indag. Math.* 43 (1981), 39–52, 53–66.
124. E. M. De Jager, *Applications of Distributions in Mathematical Physics*, Amsterdam, Math. Centre Tract 10, 1964.
125. E. M. De Jager, Theory of distributions, in E. Roubine (Ed.), *Mathematics Applied to Physics*, Springer-Verlag, N.Y., 1970.
126. P. Deligne, Formes modulaires et représentations l-adiques, *Sém. Bourbaki, 21* (1969), exp. 335.
127. P. Deligne, *La conjecture de Weil, Inst. des Hautes Études Sci., Pub. Math., 53* (1974), 273–307.
128. B. N. Delone and S. S. Ryskov, Extremal problems in the theory of positive quadratic forms, *Proc. Steklov Inst. Math., 112* (1971), 211–231.
129. J. Delsarte, *Oeuvres*, Éditions du Centre National de la Recherche Scientifique, Paris, 1971.
130. J.-M. Deshouillers and H. Iwaniec, Kloosterman sums and Fourier coefficients of cusp forms, *Inv. Math, 70* (1982/83), 219–288.
131. C. M. Dewitt, L'intégrale fonctionnelle de Feynman. Une introduction, *Ann. Inst. H. Poincaré, XI* (1969), 153–206.
132. P. Diaconis, *Group Representations in Probability and Statistics*, Inst. of Math. Statistics, Hayward, CA, 1988.
133. F. Diamond and J. Shurman, *A First Course in Modular Forms,* Springer, NY, 2005.
134. J. Dieudonné, *Treatise on Analysis*, Vols. I-VI, Academic, N.Y., 1969–1978.
135. P. A. M. Dirac, The physical interpretation of quantum dynamics, *Proc. Royal Soc., London, Sect. A, 113* (1926–1927), 621–641.
136. G. Doetsch, *Introduction to the Theory and Application of the Laplace Transform,* Springer-Verlag, N.Y., 1974.
137. W. F. Donoghue, *Distributions and Fourier Transforms*, Academic, N.Y., 1969.
138. J. S. Dowker, Quantum mechanics on group space and Huygen's principle, *Ann. of Physics, 62* (1971), 361–382.
139. J. S. Dowker, Quantum mechanics and field theory on multiply connected and on homogeneous spaces, *J. Phys. A., 5* (1972), 936–943.
140. J. R. Driscoll and D. M. Healy, Jr., Computing Fourier transforms and convolutions on the 2-sphere, *Adv. in Appl. Math.*, 15 (1994), 202–250).
141. R. M. Dudley, Lorentz-invariant Markov processes in relativistic phase space, *Arkiv for Mat., 6* (1965), 241–268.
142. R. M. Dudley, Recession of some relativistic Markov processes, *Rocky Mt. J. of Math., 4* (1974), 401–406.
143. J. J. Duistermaat, J. A. C. Kolk, and V. S. Varadarajan, Spectra of compact locally symmetric manifolds of negative curvature, *Inv. Math., 52* (1979), 27–93.
144. W. Duke, Hyperbolic distribution problems and half-integral weight Maass forms, *Invent. Math., 92* (1988), 73–90.
145. N. Dunford and J. T. Schwartz, *Linear Operators, Vols. I, II, III,* Wiley-Interscience, N.Y., 1958, 1963, 1971.

146. P. Du Val, *Elliptic Functions and Elliptic Curves*, Cambridge U. Press, Cambridge, UK, 1973.
147. H. Dym and H. P. McKean, *Fourier Series and Integrals*, Academic, N.Y., 1972.
148. F. Dyson, *Symmetry Groups in Nuclear and Particle Physics*, Benjamin, NY, 1966.
149. H. Edwards, *Riemann's Zeta Function,* Academic, N.Y., 1974.
150. M. Eichler, *Introduction to the Theory of Algebraic Numbers and Functions*, Academic, NY, 1966.
151. M. Eichler, *The basis problem for modular forms and the traces of the Hecke operators, Lecture Notes in Math., 320*, Springer-Verlag, NY, 1973, 75–151.
152. M. Eichler, Modular correspondences and their representations, *J. Indian Math. Soc., 20* (1955), 161–206.
153. M. Eichler, Eine Verallgemeinerung der Abelschen Integrale, *Math. Z., 67* (1957), 267–298.
154. M. Einsiedler, E. Lindenstrauss, P. Michel, and A. Venkatesh, The distribution of closed geodesics on the modular surface and Duke's theorem, preprint.
155. J. Elstrodt, Die Resolvente zum Eigenwertproblem der automorphen Formen in der hyperbolischen Ebene I, II, III. *Math. Ann., 203* (1973), 295–330; *Math. Z., 132* (1973), 99–134; *Math. Ann., 208* (1974), 99–132.
156. J. Elstrodt, Die Selbergsche Spurformel für Kompakte Riemannsche Flächen, *Jahresber. d. Dt. Math.-Verein, 83* (1981), 45–77.
157. J. Elstrodt, F. Grunewald, and J. Mennicke, On the group $PSL(\mathbb{Z}[i])$, J.V. Armitage (Ed.), *Journées Arithmètiques 1980 Exeter, LMS Lecture Notes*, Cambridge U. Press, Cambridge, UK, 1982.
158. J. Elstrodt, F. Grunewald, and J. Mennicke, Discontinuous groups on 3-dimensional hyperbolic space: Analytical theory and arithmetic applications, *Russian Math. Surveys, 38* (1983), 137–168.
159. J. Elstrodt, F. Grunewald, and J. Mennicke, $PSL(2)$ over imaginary quadratic integers, *Astérisque, 94* (1983), 43–60.
160. J. Elstrodt, F. Grunewald, and J. Mennicke, *Groups Acting on Hyperbolic Space: Harmonic Analysis and Number Theory*, Springer, Berlin, 1998.
161. O. Emersleben, Die elektrostatische Gitterenergie endlicher Stücke heteropolarer Kristalle, *Z. Physik Chem., 199* (1952), 170–190.
162. O. Emersleben, Über das Restglied der Gitterenergieentwicklung neutraler Ionengitter, *Math. Nach., 9* (1953), 221–234.
163. P. Epstein, Zur Theorie allgemeiner Zetafunktionen, I, II, *Math. Ann., 56, 63* (1903, 1907), 614–644, 205–216.
164. A. Erdélyi et al., *Higher Transcendental Functions, Vols. I, II, III,* McGraw-Hill, N.Y., 1953–1955.
165. A. Erdélyi et al., *Integral Transforms, Vols. I, II*, McGraw-Hill, N.Y., 1954.
166. B. Ernst, *The Magic Mirror of M.C. Escher*, Ballantine, N.Y., 1976.
167. R. Evans, Spherical functions for finite upper half planes with characteristic 2, *Finite Fields Applics, 3* (1995), 376–394.
168. R. Evans, Character sums as orthogonal eigenfunctions of adjacency operators for Cayley graphs, *Contemp. Math., 168* (1994), A.M.S., Providence, RI, pp. 33–50.
169. P. P. Ewald, Die Berechnung optischer und elektrostatischer Gitterpotentiale, *Ann. d. Phys., 64* (1921), 253–287.
170. H. Eyring, J. Walter, and G. Kimball, *Quantum Chemistry*, Wiley, NY, 1944.
171. L. D. Faddeev, Expansion in eigenfunctions of the Laplace operator on the fundamental domain of a discrete group on the Lobacevskii plane, *Trans. Moscow Math. Soc., 17* (1967), 357–386.
172. K. Falconer, *Fractal Geometry*, Wiley, Chichester, 1990.
173. J. Faraut, Dispersion d'une mesure de probabilité sur $SL(2,\mathbb{R})$ biinvariante par $SO(2,\mathbb{R})$ et théorème de la limite centrale. Preprint, 1975.
174. F. A. Farris, The edge of the universe-hyperbolic wallpaper, *Math. Horizons, 8* (Sept., 2001), 16–23.

175. F. A. Farris, Forbidden symmetries, *Notices of the Amer. Math. Soc., 59, No. 10* (Nov., 2012), 1386–1390.
176. J. D. Fay, Fourier coefficients of the resolvent for a Fuchsian group, *J. fur die Reine und Angew. Math., 293* (1977), 143–203.
177. W. Feller, *An Introduction to Probability Theory and its Applications, Vols. I, II.* Wiley, N.Y., 1950, 1966.
178. R. Feynman, *Quantum Mechanics and Path Integrals*, McGraw-Hill, NY, 1965.
179. K. L. Fields, Locally minimal Epstein zeta functions, *Mathematika, 27* (1980), 17–24.
180. H. Flanders, *Differential Forms*, Academic, NY, 1963.
181. V. A. Fock, On the representation of an arbitrary function by an integral involving Legendre's functions with a complex index, *Comptes Rendus Acad. Sci. U.R.S.S., Doklady, N.S., 39* (1943), 253–256.
182. G. B. Folland, *Fourier Analysis and its Applications,* Wadsworth and Brooks/Cole, Pacific Grove, CA, 1992.
183. L. Ford, *Automorphic Functions*, Chelsea, NY, 1951.
184. H. Frasch, Die Erzeugenden der Hauptkongruenzgruppen für Primzahlstufen, *Math. Ann., 108* (1933), 229–252.
185. R. Fricke and F. Klein, *Vorlesungen über die Theorie der automorphen Functionen, I*, Teubner, Leipzig, 1897.
186. F. G. Friedlander, *Introduction to the Theory of Distributions*, Cambridge U. Press, Cambridge, UK, 1982.
187. H. Funk, Beiträge zur Theorie der Kugelfunktionen, *Math. Ann., 77* (1916), 136–152.
188. H. Furstenberg, Noncommuting random products, *Trans. Amer. Math. Soc.*, 108 (1963), 377–428.
189. H. Furstenberg and H. Kesten, Products of random matrices, *Ann. Math. Stat., 31* (1980), 459–469.
190. M. J. Gander and F. Kwok, Chladni figures and the Tacoma bridge: motivating PDE eigenvalue problems via vibrating plates, SIAM Review, 54 (2012), 573–596.
191. R. Gangolli, Spectra of discrete uniform subgroups, in W. Boothby and G. Weiss (Eds.), *Geometry and Analysis on Symmetric Spaces*, Dekker, NY, 1972, 93–117.
192. R. Gangolli, Spherical functions on semisimple Lie groups, in W. Boothby and G. Weiss (Eds.), *Geometry and Analysis on Symmetric Spaces*, Dekker, NY, 1972, 41–92.
193. R. Gangolli, Isotropic infinitely divisible measures on symmetric spaces, *Acta Math., 111* (1964), 213–246.
194. R. Gangolli, On the Plancherel formula and the Paley-Wiener theorem for spherical functions on semisimple Lie groups, *Ann. of Math., 93* (1978), 150–165.
195. R. Gangolli and G. Warner, On Selberg's trace formula, *J. Math. Soc. Japan, 27* (1975), 328–343.
196. R. Gangolli and G. Warner, Zeta functions of Selberg's type for some noncompact quotients of symmetric spaces of rank one, *Nagoya Math. J., 78* (1980), 1–44.
197. P. R. Garabedian, *Partial Differential Equations*, Wiley, N.Y., 1964.
198. M. Gardner, *Penrose Tiles to Trapdoor Ciphers*, Cambridge U. Press, 1997.
199. C. F. Gauss, *Werke*. Königlichen Gesellshaft der Wissenshaften, Göttingen, 1870–1927.
200. S. Gelbart, *Automorphic Forms on Adele Groups*, Princeton U. Press, Princeton, NJ, 1975.
201. I. M. Gelfand, Spherical functions on symmetric spaces, *Doklady Akad. Nauk. S.S.S.R., 70* (1950), 5–8.
202. I. M. Gelfand and S. V. Fomin, Geodesic flows on manifolds, *A.M.S. Translations, 2, Vol. 1*, (1955), 49–66.
203. I. M. Gelfand, M. I. Graev, and I. I. Piatetskii-Shapiro, *Representation Theory and Automorphic Functions*, Saunders, Philadelphia, PA, 1966.
204. I. M. Gelfand, M. I. Graev, and N. Ya. Vilenkin, *Generalized Functions, Vol. 5,* Academic, NY, 1966.

205. I. M. Gelfand, R. A. Minlos, and Z. Ya. Shapiro, *Representations of the Rotation and Lorentz Groups and Their Applications,* MacMillan, NY, 1963.
206. I. M. Gelfand and G. E. Shilov, *Generalized Functions, Vols. 1–3*, Academic, N.Y., 1964.
207. I. M. Gelfand, M. I. Graev, and N. Ya. Vilenkin, *Generalized Functions, Vol. 5*, Academic, NY, 1966.
208. M. E. Gertsenshtein and V. B. Vasil'ev, Waveguides with random inhomogeneities and Brownian motion in the Lobachevsky plane, *Theory of Probability and its Applications, 4* (1959), 391–398.
209. R. K. Getoor, Infinitely divisible probabilities on the hyperbolic plane, *Pacific J. Math., 11* (1961), 1287–1308.
210. E. Ghys, Knots and dynamics, 2006 *Intl. Cong. of Math. Proc., I,* 247–277, 2006.
211. E. Ghys and J. Leys, Lorenz and Modular Flows: A Visual Introduction, *Feature Column, AMS*, www.ams.org/featurecolumn/archive/lorenz.html.
212. F. Gilbert, Inverse problems for the earth's normal modes, in W.H. Reid (Ed.), *Mathematical Problems in the Geophysical Sciences, Vol. I,* Amer. Math. Soc., Providence, RI, 1971, 107–128.
213. R. Godement, Introduction aux travaux de A. Selberg, *Sém. Bourbaki, exp. 144*, Paris, 1957.
214. R. Godement, La formule des traces de Selberg considerée comme source de problèmes mathématiques, *Sém. Bourbaki, exp. 244*, Paris, 1962.
215. R. Godement, Introduction à la théorie de Langlands, *Sém. Bourbaki, exp. 321* (1966/67), Benjamin, NY, 1968.
216. R. Godement, A theory of spherical functions, *Trans. American Math. Soc., 73* (1952), 496–556.
217. D. Goldfeld, Explicit formulae as trace formulae, in *Number theory, trace formulas and discrete groups (Oslo, 1987)*, 281–288, Academic Press, Boston, MA, 1989.
218. D. Goldfeld, J. Hoffstein, and S. J. Patterson, Mean values of certain elliptic L-series, in N. Koblitz (Ed.), *Number Theory Related to Fermat's Last Theorem*, Birkhäuser, Boston, 1982.
219. D. Goldfeld and J. Hundley, *Automorphic Representations and L-Functions for the General Linear Group, Vol. I*, Cambridge U. Press, Cambridge, UK, 2011.
220. D. Goldfeld and D. Husemoller, On $L^2\left(SL_2(\mathbb{Z})\right)\backslash H)$, preprint.
221. D. Goldfeld and P. Sarnak, Sums of Kloosterman sums, *Inv. Math., 71* (1983), 243–250.
222. D. Goldfeld and C. Viola, Mean values of L-functions associated to elliptic, Fermat and other curves at the centre of the critical strip, *J. Number Theory, 11* (1979), 305–320.
223. L. J. Goldstein, *Analytic Number Theory,* Prentice-Hall, Englewood Cliffs, NJ, 1971.
224. L. J. Goldstein, Dedekind sums for a Fuchsian group I, II, *Nagoya Math. J., 50* (1973), 21–47; *53* (1974), 171–187, 235–237.
225. L. J. Goldstein, On a conjecture of Hecke conceming elementary class number formulas, *Manuscripta Math. 9* (1973), 245–305.
226. L. J. Goldstein and M. Razar, A generalization of Dirichlet's class number formula, *Duke Math. J., 43* (1976), 349–358.
227. C. S. Gordon, D. L. Webb, and S. Wolpert, Isospectral plane domains and surfaces via Riemannian orbifolds, *Inv. Math., 110* (1992), 1–22.
228. I. S. Gradshteyn and I. M. Ryzhik, Tables of Integrals, Series and Products, Academic, N.Y., 1965.
229. A. Granville and Z. Rudnick, *Equidistribution in Number Theory, An Introduction, Proc. NATO Adv. Study Inst., Montréal, Canada, 2005*, Springer, Dordrecht, The Netherlands, 2007.
230. I. Grattan-Guinness and J. R. Ravetz, *Joseph Fourier*, M.I.T. Press, Cambridge, MA, 1972.
231. D. Greenspan, *Discrete Models*, Addison-Wesley, Reading, MA, 1973.
232. P.R. Griffiths, Fourier transform infrared spectrometry, *Science, 222* (1983), 297–302.
233. E. Grosswald, *Topics from the Theory of Numbers*, MacMillan, NY, 1966.
234. F. A. Grünbaum, L. Longhi, and M. Perlstadt, Differential operators commuting with finite convolution integral operators: some nonabelian examples, *SIAM J. Appl. Math., 42* (1982), 941–955.

235. H. Guggenheimer, *Differential Geometry,* Dover, NY, 1977.
236. R. C. Gunning, *Lectures on Modular Forms*, Princeton U. Press, Princeton, NJ, 1962.
237. R. Gupta, *Fields of division points of elliptic curves related to Coates-Wiles*, Ph.D. Thesis, M.I.T., 1983.
238. D. Gurarie, *Symmetries and Laplacians. Introduction to Harmonic Analysis, Group Representations and Applications*, Elsevier, Amsterdam, The Netherlands, 1992.
239. M. Gutzwiller, Stochastic behavior in quantum scattering, *Physica D, 7* (1983), 341–355.
240. M. Gutzwiller, The quantization of a classically ergodic system, *Physica D, 5* (1982), 183–207.
241. M. Gutzwiller, Classical quantization of a Hamiltonian with ergodic behavior, *Physical Review Letters, 45* (1980), 150–153.
242. M. Gutzwiller, *Chaos in Classical and Quantum Mechanics*, Springer-Verlag, NY, 1990.
243. R. K. Guy, *Reviews in Number Theory, 1973–1983*, Amer. Math. Soc., Providence, RI, 1984.
244. H. Haas, *Numerische Berechnung der Eigenwerte der Differentialgleichung* $y^2 \Delta u + \lambda u = 0$ *für ein unendliches Gehiet im* $\mathbb{R}^2$, Universität Heidelberg, Diplomarbeit, 1977.
245. Y. Hamahata, A note on modular forms on finite upper half planes, in *Lecture Notes in Computer Science, 4547*, Springer-Verlag, Berlin, 2007, pp. 18–24.
246. M. Hamermesh, *Group Theory and its Applications to Physical Problems*, Addison-Wesley, Reading, MA, 1962.
247. R. W. Hamming, *Digital Filters,* Prentice-Hall, Englewood Cliffs, N.J., 1977.
248. L. Hantsch and W. von Waldenfels, A random walk on the general linear group related to a problem of atomic physics, *Lecture Notes in Math., Vol. 706*, Springer-Verlag, NY, 1979, 131–143.
249. G. H. Hardy, Ramanujan, Chelsea, NY, 1940.
250. G. H. Hardy, *Collected Papers, Vols. 1–7,* Oxford U. Press, Oxford, UK, 1966–1979.
251. G. H. Hardy and E. M. Wright, *An Introduction to the Theory of Numbers,* Oxford U. Press, London, 1956.
252. Harish-Chandra, *Automorphic Forms on Semi-Simple Lie Groups, Lecture Notes in Math., 62,* Springer-Verlag, N.Y., 1968.
253. Harish-Chandra, Spherical functions on a semisimple Lie group I, II, *Amer. J. Math., 80* (1953), 241–310, 553–613.
254. Harish-Chandra, Representations of semisimple Lie groups I, II, III, *Trans. Amer. Math. Soc., 75* (1953), 185–243; *76* (1954), 26–65, 234–253.
255. D. C. Harris and M. D. Bertolucci, *Symmetry and Spectroscopy,* Oxford U. Press, NY, 1978.
256. C. B. Haselgrove and J. C. P. Miller, *Tables of the Riemann Zeta Function, Royal Society Math. Tables, 6*, Cambridge U. Press, Cambridge, UK, 1963.
257. D. R. Heath-Brown and S. J. Patterson, The distribution of Kummer sums at prime arguments, *J. Reine Angew. Math., 310* (1979), 111–130.
258. E. Hecke, *Mathematische Werke*, Vandenhoeck und Ruprecht, Göttingen, 1970.
259. E. Hecke, *Lectures on Dirichlet Series, Modular Functions and Quadratic Forms*, given at the Institute for Advanced Study in 1938, B. Schoeneberg and W. Maak (Eds.). Vandenhoeck und Ruprecht, Göttingen, Germany, 1983.
260. E. Hecke, *Lectures on the Theory of Algebraic Numbers,* Springer-Verlag, NY, 1981.
261. D. Hejhal, *The Selberg Trace Formula for PSL*$(2,\mathbb{R})$*, Vols. I, II, Lecture Notes in Math., 548, 1001*, Springer-Verlag, N.Y., 1976, 1983.
262. D. Hejhal, The Selberg trace formula and the Riemann zeta function, *Duke Math. J., 43* (1976), 441–482.
263. D. Hejhal, Some observations concerning eigenvalues of the Laplacian and Dirichlet L-Series, in H. Halberstam and C. Hooley (Eds.), *Recent Progress in Analytic Number Theory*, Academic, London, UK, 1981, pp. 95–110.
264. D. Hejhal, Some Dirichlet series whose poles are related to cusp forms, Pts. I, II, *Technical Reports, 1981–14, 1982–13, Chalmers University of Tech.*, Sweden, 1981, 1982.
265. D. A. Hejhal, A Note on the Voronoi summation formula, *Monatshefte für Math., 87* (1979), 1–14.

266. D. A. Hejhal, On the calculation of Maass cusp forms, in *Hyperbolic Geometry and Applications in Quantum Chaos and Cosmology*, J. Bolte and F. Steiner, Eds., Cambridge U. Press, Cambridge, UK, 2012, pp. 175–186.
267. D. A. Hejhal, Sur certaines séries de Dirichlet associées aux géodésiques fermées d'une surface de Riemann compacte, *C.R. Acad. Sci. Paris, 294* (1982), 273–276.
268. D. A. Hejhal, Sur quelques propriétés asymptotiques des périodes hyperboliques et des invariants algebriques d'un sous-groupe discret de $PSL(2,\mathbb{R})$, *C.R. Acad. Sci. Paris, 294* (1982), 509–512.
269. D. A. Hejhal, Quelques examples de séries de Dirichlet dont les pôles ont un rapport étroit avec les valeurs propres de l'opérateur de Laplace-Beltrami hyperbolique, *C.R. Acad. Sci. Paris, 294* (1982), 637–640.
270. D. A. Hejhal, Some Dirichlet series with coefficients related to periods of automorphic eigenforms [parts 1–2], *Proc. Japan Acad. A., 58, 59* (1982, 1983) 413–417, 335–338.
271. D. A. Hejhal, A continuity method for spectral theory on Fuchsian groups, in R.A. Rankin (Ed.), *Modular Forms*, Horwood, Chichester (distrib. Wiley), 1984, 107–140.
272. D. A. Hejhal, A classical approach to a well known spectral correspondence on quaternion groups, *Lecture Notes in Math., Vol. 1135*, Springer-Verlag, Berlin, 1985, pp. 127–190.
273. D. A. Hejhal, *Eigenvalues of the Laplacian for Hecke Triangle Groups, Memoirs of the Amer. Math. Soc., 469*, AMS, Providence, RI, 1992.
274. D. A. Hejhal and B. Rackner, On the topography of Maass wave forms for $PSL(2,\mathbb{Z})$, *Experimental Math., 1* (1992), 275–305.
275. D. A. Hejhal and A. Strömbergsson, Chaos and Maass waveforms of CM-type, Invited papers dedicated to Martin C. Gutzwiller, Part IV. *Found. Phys. 31* (2001), no. 3, 519–533.
276. S. Helgason, Lie Groups and Symmetric Spaces, in C. M. DeWitt and J. A. Wheeler (Eds.), *Battelle Recontres,* Benjamin, NY, 1968, 1–71.
277. S. Helgason, *Differential Geometry and Symmetric Spaces*, Academic, NY, 1962.
278. S. Helgason, *Differential Geometry, Lie Groups, and Symmetric Spaces,* Academic, NY, 1978.
279. S. Helgason, *The Radon Transform*, Birkhä user, Boston, MA, 1980.
280. S. Helgason, *Topics in Harmonic Analysis on Homogeneous Spaces*, Birkhäuser, Boston, MA, 1981.
281. S. Helgason, Functions on symmetric spaces, *Proc. Symp. Pure Math., 26,* A.M.S., Providence, 1973, 101–146.
282. S. Helgason, *Analysis on Lie Groups and Homogeneous Spaces, A.M.S. Regional Conference, 14*, A.M.S., Providence, RI, 1971 (corrected, 1977).
283. S. Helgason, An analogue of the Paley-Wiener theorem for the Fourier transform on certain symmetric spaces, *Math. Ann., 165* (1966), 297–308.
284. S. Helgason, A duality for symmetric spaces with applications to group representations, III. Tangent space analysis, *Adv. in Math., 36* (1980), 297–323.
285. J. Helszajn, *Passive and Active Microwave Circuits*, Wiley, NY, 1978.
286. J. W. Helton, Non-Euclidean functional analysis and electronics, *Bull. Amer. Math. Soc., 7* (1982), 1–64.
287. J. W. Helton, The distance of a function to H^∞ in the Poincaré metric. Electrical power transfer, J*ournal of Functional Analysis, 38* (1980), 273–314.
288. J. W. Helton, A simple test to determine gain bandwidth limitations. *Proc. I.E.E.E. Internatl. Conf. on Circuits and Systems*, 1977.
289. P. Henrici, *Applied and Computational Complex Analysis, Vols. I, II*, Wiley, N.Y., 1974, 1977.
290. G. T. Herman (Ed.), *Image Reconstruction from Projections,* Springer-Verlag, NY, 1979.
291. R. Hermann, *Lie Groups for Physicists*, Benjamin, NY, 1966.
292. R. Hermann, *Fourier Analysis on Groups and Partial Wave Analysis*, Benjamin, NY, 1969.
293. Hewlett-Packard Co., *Microwave Theory and Measurements*, Prentice-Hall, Englewood Cliffs, NJ, 1963.
294. H. Heyer, An application of the method of moments to the central limit theorem, *Lecture Notes in Math., Vol. 861*, Springer-Verlag, NY, 1981, 65–73.

295. H. Hijikata, A. Pizer, and T. Shemanske, The basis problem for modular forms on $\Gamma_0(N)$, *Proc. Japan Acad., 56* (1980), 280–284.
296. E. Hilb and M. Riesz, Neuere Untersuchungen über Trigonometrische Reihen, *Enzycl. der Math. Wissen.*, Teubner, Leipzig, 1916.
297. D. Hilbert and S. Cohn-Vossen, *Geometry and the Imagination,* Chelsea, N.Y., 1952.
298. K. Højendahl, Studies in the properties of crystals III, On an elementary method of calculating Madelung's constant, *Danske Vid. Selske, Math.-fys. Medd., 16* (1982), 133–154.
299. R. Holowinsky and K. Soundararajan, Mass equidistribution of Hecke eigenfunctions, *Ann. of Math. (2), 172* (2010), 1517–1528.
300. R. T. Hoobler and H. L. Resnikoff, Normal connections for automorphic embeddings. Preprint.
301. C. Hooley, On the distribution of the roots of polynomial congruences, *Mathematika, 11* (1964), 39–49.
302. S. Hoory, N. Linial, and A. Wigderson, Expander graphs and their applications, *Bull. A.M.S.*, 43 (2006), 439–561.
303. L. Hörmander, On the division of distributions by polynomials, *Arkiv. Matem., 3* (1958), 555–568.
304. M. D. Horton and D. Newland, The Contest between the Kernels in the Selberg Trace Formula for the $(q+1)$-regular Tree, in *The Ubiquitous Heat Kernel,* J. Jorgenson and L. Walling (Eds.), *Contemporary Mathematics, Volume 398* (2006), pages 265–294.
305. J. Horvath, *An introduction to distributions*, *Amer. Math. Monthly, 77* (1970), 227–240.
306. H. Huber, Über eine neue Klasse automorphe Funktionen und eine Gitterpunkt-problem in der hyperbolischen Ebene, *Comment. Math. Helvet., 30* (1965), 20–62.
307. H. Huber, Zur analytischen Theorie hyperbolischer Raumformen und Bewegungsgruppen, *Math. Ann., 138* (1959), 1–26; II, *Math. Ann., 142* (1961), 385–398; Nachtrag zu II, *Math. Ann., 143* (1961), 463–464.
308. N. Hurt, *Geometric Quantization in Action*, D. Reidel, Amsterdam, 1983.
309. N. Hurt, Propagators in quantum mechanics on multiply connected spaces, *Lecture Notes in Physics, 50*, Springer-Verlag, NY, 1976, 182–192.
310. N. Hurt and R. Hermann, *Quantum Statistical Mechanics and Lie Group Harmonic Analysis*, Part A, Math. Sci. Press, Brookline, MA 1982.
311. A. Hurwitz, Über algebraische Korrespondenzen und das allgemeine Korrespondenzprinzip, *Math. Werke, I*, Eidgenössischen Technischen Hochschule, Basel, 1932, 163–188.
312. A. Hurwitz and R. Courant, *Funktionentheorie*, Springer-Verlag, NY, 1964.
313. J.-I. Igusa, *Theta Functions*, Springer-Verlag, NY, 1974.
314. Y. Ihara, Hecke polynomials as congruence ζ-functions in elliptic modular case, *Ann. of Math., (2) 85* (1967), 267–295.
315. K. Imai (Ota), Generalization of Hecke's correspondence to Siegel modular forms, *Amer. J. Math., 102* (1980), 903–936.
316. L. Infeld and T. E. Hull, The Factorization Method, *Rev. Mod. Phys. 23,* (1951), 21–68.
317. K. Ireland and M. Rosen, *A Classical Introduction to Modern Number Theory*, Springer-Verlag, NY, 1982.
318. H. Iwaniec, *Introduction to the Spectral Theory of Automorphic Forms*, Revista Matemática Iberoamericana, 1995.
319. H. Iwaniec, *Topics in Classical Automorphic Forms*, Amer. Math. Soc., Providence, RI, 1997.
320. H. Iwaniec and E. Kowalski, *Analytic Number Theory*, Amer. Math. Soc., Providence, RI, 2004.
321. K. Iwasawa, *Lectures on p-adic L-Functions*, Princeton U. Press, Princeton, N.J., 1972.
322. C. G. J. Jacobi, *Gesammelte Werke, Vol. I*, Chelsea, N.Y., 1969.
323. H. Jacquet and R. P. Langlands, *Automorphic Forms on GL*(2)*, Lecture Notes in Math, 114*, Springer-Verlag, NY, 1970.
324. H. Jacquet and J. Shalika, A non-vanishing theorem for zeta functions of $GL(2)$, *Inv.Math., 38* (1976), 1–16.

325. D. Jakobson, Quantum unique ergodicity for Eisenstein series on $PSL2(\mathbb{Z})\backslash PSL2(\mathbb{R})$, *Ann. Inst. Fourier (Grenoble), 44* (1994), 1477–1504.
326. T. Janssen, *Crystallographic Groups,* North Holland-Elsevier, N.Y., 1973.
327. H. and B. Jeffreys, *Methods of Mathematical Physics*, Cambridge U. Press, Cambridge, UK, 1972.
328. G. M. Jenkins and D. G. Watts, *Spectral Analysis and its Applications,* Holden-Day, San Francisco, 1968.
329. N. Jochnowitz, The index of the Hecke ring, T_k, in the ring of integers of $T_k \otimes \mathbb{Q}$, *Duke Math. J., 46* (1979), 861–869.
330. M. Kac, *Selected Papers*, M.I.T. Press, Cambridge, MA, 1979.
331. M. Kac, The search for the meaning of independence, in J. Gani, (Ed.), *The Making of Statisticians,* Springer-Verlag, N.Y., 1982, 62–72.
332. M. Kac, Can you hear the shape of a drum?, *Amer. Math. Monthly, 73* (1966), 1–23.
333. M. Kac, *Enigmas of Chance: An Autobiography*, U. of California Press, Berkeley, CA, 1985.
334. V. G. Kac, An elucidation of "infinite-dimensional algebras,. . . and the very strange formula." $E_8^{(1)}$ and the cube root of the modular invariant j, *Advances in Math., 35* (1981), 264–273.
335. V. G. Kac, Infinite-dimensional algebras, Dedekind's η-function, classical Möbius function and the very strange formula, *Advances in Math., 30* (1978), 85–136.
336. E. R. Kanasewich, *Time Sequence Analysis in Geophysics,* U. of Alberta Press, Alberta, Canada, 1975.
337. F. I. Karpelevich, V. N. Tutubalin, and M. G. Shur, Limit theorems for the compositions of distributions in the Lobachevsky plane and space, *Theory of Probability and its Applications, 4* (1959), 399–402.
338. M. Kashiwara, A. Kowata, K. Minemura, K. Okamoto, T. Oshima, and M. Tanaka, Eigenfunctions of invariant differential operators on a symmetric space. *Ann. Math., 107* (1978), 1–39.
339. A. Kasman, *Glimpses of Soliton Theory: The Algebra and Geometry of Nonlinear PDEs, Amer. Math. Soc., Student Mathematical Library, Vol. 54*, Providence, RI, 2010.
340. K. Katayama, Kronecker's limit formulas and their applications, *J. Fac. Sci. U. Tokyo, 13* (1966), 1–44.
341. S. Kato, *Dynamics of the Upper Atmosphere,* Reidel, Amsterdam, Holland, 1980.
342. S. Katok, *Fuchsian Groups,* U. of Chicago Press, Chicago, IL, 1992.
343. S. Katok and I. Ugarcovici, Symbolic dynamics for the modular surface and beyond, *Bull. Amer. Math. Soc., 44, No. 1* (Jan., 2007), 87–132.
344. N. Katz, An overview of Deligne's proof of the Riemann hypothesis for varieties over function fields (Hilbert's Problem 8), *Proc. Symp. in Pure Math., Vol. 28*, American Math. Soc., Providence, RI, 1976, 286–288.
345. N. Katz, Sommes exponentielles, *Astérisque, 79* (1980), available from Amer. Math. Soc. in U.S.
346. N. Katz, Estimates for Soto-Andrade sums, *J. Reine Angew. Math., 438* (1993), 143–161.
347. Y. Katznelson, *An Introduction to Harmonic Analysis,* Wiley, N.Y., 1968.
348. J. P. Keener, *Principles of Applied Mathematics: Transformation and Approximation*, Addison-Wesley, Redwood City, CA, 1988.
349. J. B. Keller and G. C. Papanicolaou, Stochastic differential equations with applications to random harmonic oscillators and wave propagation in random media, *SIAM J. Appl. Math., 21* (1971), 287–305.
350. L. J. P. Kilford, *Modular Forms: A Classical and Computational Introduction*, Imperial College Press, London, UK, 2008.
351. H. H. Kim, Functoriality for the exterior square of GL4 and the symmetric fourth of GL2, *Journal of the Amer. Math. Soc., 16* (2003), no. 1, 139–183, with appendix 1 by D. Ramakrishnan, and appendix 2 by Kim and P. Sarnak.
352. J. F. C. Kingman, Random walks with spherical symmetry, *Acta Math., 109* (1963), 11–53.
353. A. B. Kirillov, Elements of the Theory of Representations, Springer-Verlag, NY, 1976.

354. K. Kirsten and F. L. Williams (Eds.), *A Window into Zeta and Modular Physics, M.S.R.I. Pub., Vol. 57,* Cambridge U. Press, Cambridge, UK, 2010.
355. F. Klein and R. Fricke, *Vorlesungen über die Theorie der elliptischen Modulfunktionen, Vol. I*, Johnson Reprint Corp., NY, 1966.
356. A. Knapp, *Representation Theory of Real Semisimple Groups: an Overview Based on Examples*, Princeton University Press, Princeton, NJ, 1986.
357. M. Knopp, *Modular Functions in Analytic Number Theory*, Markham, Chicago, IL, 1970.
358. S. Kobayashi, *Hyperbolic Manifolds and Holomorphic Mappings*, Dekker, NY, 1970.
359. N. Koblitz, *Introduction to Elliptic Curves and Modular Forms*, Springer-Verlag, NY, 1993.
360. N. Koblitz, *A Course in Number Theory and Cryptography*, Springer-Verlag, NY, 1987.
361. N. Koblitz, The Uneasy Relationship Between Mathematics and Cryptography, *Notices Amer. Math. Soc., Vol. 54, No. 8*, Sept., 2007, 972–979.
362. N. Koblitz and A. Menezes, The Brave New World of Bodacious Assumptions in Cryptography, *Notices Amer. Math. Soc., Vol. 57, No. 3*, March, 2010, 357–365.
363. D. R. Kohel and H. A. Verrill, Fundamental domains for Shimura curves, *J. de Théorie des Nombres de Bordeaux, 15* (2003), 205–222.
364. A. N. Kolmogorov, Une série de Fourier-Lebesgue divergente partout, *C.R. Acad. Sci. Paris, Ser. A.-B, 183* (1926), 1327–1328.
365. A. N. Kolmogorov, *Foundations of the Theory of Probability,* Chelsea, N.Y., 1950.
366. A. N. Kolmogorov and S. V. Fomin, *Introductory Real Analysis*, Dover, N.Y., 1975.
367. M. J. Kontorovich and N. N. Lebedev, *J. Exper. Theor. Phys. U.S.S.R. 8* (1938), 1192–1206.
368. M. J. Kontorovich and N. N. Lebedev, *Acad. Sci. U.S.S.R., J. Physics, 1* (1939), 229–241.
369. M. J. Kontorovich and N. N. Lebedev, *J. Exper. Theor. Phys. U.S.S.R., 9* (1939), 729–741.
370. A. Koranyi, A survey of harmonic functions on symmetric spaces, *Proc. Symp. Pure Math., 35*, Amer. Math. Soc., Providence, RI, 1979, 323–344.
371. J. Korevaar, *Mathematical Methods*, Academic, N.Y., 1968.
372. T. W. Körner, *Fourier Analysis*, Cambridge U. Press, Cambridge, UK, 1988.
373. A. Krieg, *Hecke Algebras, Memoirs of Amer. Math. Soc., 435*, Providence, 1990.
374. L. Kronecker, *Werke*, Leipzig, 1929.
375. T. Kubota, *Elementary Theory of Eisenstein Series*, Wiley, NY, 1973.
376. T. Kubota, Some results concerning reciprocity law and real analytic automorphic functions, *Proc. Symp. Pure Math., Vol. 20*, Amer. Math. Soc., Providence, RI, 1970, 382–395.
377. S. Kudla, Relations between automorphic forms produced by theta functions, *Lecture Notes in Math., 627,* Springer-Verlag, NY, 1977, 277–285.
378. S. Kudla, Theta functions and Hilbert modular forms, *Nagoya Math. J., 69* (1978), 97–106.
379. S. Kudla and J. Millson, Harmonic differentials and closed geodesics on Riemann surfaces, *Inv. Math., 54* (1979), 193–211.
380. A. Kurihara, On some examples of equations defining Shimura curves and the Mumford uniformization, *J. Fac. Sci. Univ. Tokyo, 25* (1979), 277–301.
381. N. V. Kuznetsov, Petersson's conjecture for cusp forms of weight zero and Linnik's conjecture. Sums of Kloosterman sums, *Math. U.S.S.R. Sbornik, 39* (1981), 299–342.
382. G. Lachaud, Analyse spectrale des formes automorphes et séries d'Eisenstein, *Inv. Math., 46* (1978), 39–79.
383. J. Lagarias and A.M. Odlyzko, On computing Artin L-functions in the critical strip, *Math. Comp., 33* (1979), 1081–1095.
384. I. Lakatos, *Proofs and Refutations*, Cambridge U. Press, Cambridge, 1976.
385. A. L. Lance, *Introduction to Microwave Theory and Measurements,* McGraw-Hill, NY, 1969.
386. S. Lang, *Analysis I*, Addison-Wesley, Reading, Mass., 1968. (the revised version is *Undergraduate Analysis*, Springer-Verlag, N.Y., 1983, 1997).
387. S. Lang, *Real Analysis*, Addison-Wesley, Reading, Mass., 1983.
388. S. Lang, *Algebraic Number Theory*, Addison-Wesley, Reading, Mass., 1968.
389. S. Lang, $SL(2,\mathbb{R})$, Addison-Wesley, Reading, MA, 1975.
390. S. Lang, *Elliptic Functions*, Addison-Wesley, Reading, MA, 1973.

391. S. Lang, *Introduction to Modular Forms*, Springer-Verlag, NY, 1976.
392. S. Lang, Some history of the Shimura-Taniyama conjectures, *Notices Amer. Math. Soc., 42* (1995), 1301–1307.
393. R. P. Langlands, *Eisenstein Series. Lecture Notes in Math., 544*, Springer-Verlag, NY, 1976.
394. R. P. Langlands, *Base Change for GL(2)*, Princeton U. Press, Princeton, NJ, 1980.
395. E. M. Lapid, *Introductory notes on the trace formula*, http://www.math.huji.ac.il/\char126\relaxerezla/papers/ch5.pdf
396. P. S. De Laplace, Théorie des attractions des sphéroides et de la figure des planètes, *Mém. de l' Acad.* (1782), Paris, 1785.
397. K. Lauter, The advantages of elliptic curve cryptography for wireless security, *IEEE Wireless Communications*, Feb., 2004, 2–7
398. A. F. Lavrik, The principle of the theory of nonstandard functional equations for the Dirichlet functions, *Proc. Steklov Inst. Math., 132* (1973), 77–85.
399. A. F. Lavrik, On functional equations of Dirichlet functions, *Math. U.S.S.R. Izvestija, 31* (1967), 421–432.
400. P. Lax and R. Phillips, *Scattering Theory for Automorphic Functions*, Princeton University Press, Princeton, NJ, 1976.
401. N. N. Lebedev, *Special Functions and their Applications*, Dover, N.Y., 1972.
402. N. N. Lebedev, Sur une formule d'inversion, *Dokl. Akad. Nauk. S.S.S.R., 52* (1946), 655–658.
403. P. Le Corbeiller, *Crystals and the Future of Physics, The World of Mathematics, Vol. II*, (J.R. Newman, Ed.). Simon and Schuster, NY, 1956, p. 876.
404. A. M. Legendre, Recherches sur l'attraction des sphéroides homogènes, *Mém. math, phys. prés. à l'Acad. Sci. par. divers savantes, 10* (1785), 411–434.
405. A. M. Legendre, Recherches sur la figure des planètes, *Mém. math. phys., Reg. de l' Acad. Sci.* (1784, 1787), 370–389.
406. D. H. Lehmer, Note on the distribution of Ramanujan's tau function, *Math. Comp., 24* (1970), 741–743.
407. D. H. Lehmer, Ramanujan's function $\tau(n)$, *Duke Math. J., 10* (1943), 483–492.
408. D. H. Lehmer, Some functions of Ramanujan, *Math. Student, 27* (1959), 105–116.
409. D. H. Lehmer, Note on the distribution of Ramanujan's tau function, *Math. Comp., 24* (1970), 741–743.
410. J. Lehner, *Discontinuous groups and automorphic functions, Math. Survey, 8*, Amer. Math. Soc., Providence, RI, 1964.
411. J. Lehner, The Picard theorems, *Math. Monthly*, 76 (1969), 1005–1012.
412. E. Le Page, Théorèmes limites pour les produits de matrices aléatoires, *Lecture Notes in Math., Vol. 928,* Springer-Verlag, NY, 1982, 258–303.
413. G. Letac, Problèmes classiques de probabilité sur un couple de Gelfand, *Lecture Notes in Math., Vol. 861*, Springer-Verlag, NY, 1981, 93–120.
414. W. Leveque, *Reviews in Number Theory*, American Math. Soc., Providence, R.I., 1974.
415. B. M. Levitan and I. S. Sargsjan, *Introduction to Spectral Theory, American Math. Soc. Transl. of Math. Monographs, 39*, Amer. Math. Soc., Providence, RI, 1975.
416. W.-C. W. Li, Hecke-Weil-Jacquet-Langlands theorem revisited, *Lecture Notes in Math., 751*. Springer-Verlag, N.Y., 206–220.
417. W.-C. W. Li, On a theorem of Hecke-Weil-Jacquet-Langlands, in H. Halberstam and C. Hooley (Eds.), *Recent Progress in Analytic Number Theory*, Academic, London, 1981, 119–152.
418. W.-C. W. Li, Newforms and functional equations, *Math. Ann.*, 212 (1975), 285–315.
419. W.-C. W. Li, *Number Theory with Applications,* World Scientific, Singapore, 1996.
420. S. Lichtenbaum, Values of zeta-functions, etale cohomology, and algebraic K-theory, *Lecture Notes in Math., Vol. 342*, Springer-Verlag, 1973, pp. 489–501.
421. M. J. Lighthill, *Introduction to Fourier Analysis and Generalized Functions,* Cambridge U. Press, Cambridge, 1978.
422. E. Lindenstrauss, Invariant measures and arithmetic quantum unique ergodicity, *Ann. of Math. (2), 163* (2006), 165–219.

423. G. Lion and M. Vergne, *The Weil Representation, Maslov Index and Theta Series*, Birkhäuser, Boston, MA, 1980.
424. A. L. Loeb, A systematic survey of cubic crystal structures, *J. Solid State Chem., 1* (1975), 237–267.
425. M. Loeve, *Probability Theory,* Van Nostrand, Princeton, N.J., 1963.
426. S. Łojasiewicz, Sur le problème de la division, *Stud. Math., 18* (1959), 87–136.
427. J. S. Lomont, *Applications of Finite Groups,* Academic, N.Y., 1959.
428. L. H. Loomis and S. Sternberg, *Advanced Calculus*, Addison-Wesley, Reading, MA, 1968.
429. K. Lonngren and A. Scott (Eds.), *Solitons in Action*, Academic, NY, 1978.
430. J.D. Louck and N. Metropolis, Hidden symmetry and the number-theoretic structure of the energy levels of a perturbed harmonic oscillator. *Adv. in Appl. Math., 1* (1980), 182–220.
431. A. K. Louis, Orthogonal function series expansions and the null space of the Radon transform. *S.I.A.M. J. Math. Analysis, 15* (1984), 621–633.
432. A. K. Louis, Optimal sampling in nuclear magnetic resonance (NMR) tomography, *J. Computer Assisted Tomography, 6* (1982), 334–340.
433. A. Lubotzky, *Discrete Groups, Expanding Graphs, and Invariant Measures*, Birkhäuser, Basel, 1994.
434. A. Lubotzky, R. Phillips, and P. Sarnak, Ramanujan graphs, *Combinatorica*, 8(1988), 261–277.
435. D. Ludwig, The Radon transform on Euclidean space, *Comm. on Pure and Appl. Math., 29* (1966), 49–81.
436. W. Luo, Z. Rudnick, and P. Sarnak, On Selberg's eigenvalue conjecture, *Geom. Funct. Anal. 5* (1995), 387–401.
437. H. Maass, *Lectures on Modular Functions of One Complex Variable,* Tata Institute of Fundamental Research, Bombay, India, 1964.
438. H. Maass, *Siegel's Modular Forms and Dirichlet Series, Lecture Notes in Math. 216*, Springer-Verlag, NY, 1971.
439. H. Maass, Über eine neue Art von nichtanalytischen automorphen Funktionen und die Bestimmung Dirichletscher Reihen durch Funktionalgleichung, *Math. Ann., 121* (1949), 141–183.
440. I. G. Macdonald, *Symmetric Functions and Hall Polynomials,* Clarendon Press, Oxford, UK, 1979.
441. I. G. Macdonald, Affine root systems and Dedekind's eta function, *Inv. Math., 15* (1972), 91–143.
442. G. Mackey, *Unitary Group Representations in Physics, Probability, and Number Theory,* Benjamin/Cummings, Reading, MA., 1978.
443. G. Mackey, *Harmonic Analysis as the Exploitation of Symmetry-A Historical Survey, Rice U. Studies, 64,* Houston, Texas, 1978.
444. G. Mackey, *The Theory of Group Representations,* U. of Chicago Press, Chicago, 1976.
445. F. J. Macwilliams and N. J. A. Sloane, *The Theory of Error-Correcting Codes,* North-Holland, Amsterdam, 1978.
446. W. Magnus, *Noneuclidean Tesselations and Their Groups*, Academic, N.Y., 1974.
447. P. Magnusson, *Transmission Lines and Wave Propagation*, Allyn and Bacon, Boston, MA, 1965.
448. J. I. Manin, Cyclotomic fields and modular curves, *Russ. Math. Surveys, 26* (1971), 7–78.
449. J. I. Manin, Parabolic points and zeta functions of modular curves, *Izv. Akad. Nauk. S.S.S.R., 6* (1972), 19–64.
450. J. I. Manin, Correspondences, motives and monoidal transformations, *Mat. Sb., 6* (1950), 439–470.
451. J. I. Manin, Modular forms and number theory, *Proc. Internatl. Cong. Math., 1978, Helsinki, Vol. I*, 177–186.
452. Y. I. Manin and A. A. Panchishkin, *Introduction to Modern Number Theory, Encyclopaedia of Mathematical Sciences, Vol. 49*, Springer, Berlin, 2005.
453. R. J. Marks, *Handbook of Fourier Analysis and Its Applications,* Oxford, NY, 2009.

454. D. K. Maslen and D. N. Rockmore, The Cooley-Tukey FFT and Group Theory, *Notices of the Amer. Math. Soc.*, 48, No. 10 (Nov., 2001), 1151–1160.
455. J. Masley and H. Montgomery, Cyclotomic fields with unique factorization, *J. Reine Angew. Math., 286* (1976), 248–256.
456. Math. Reviews Staff, *Reviews in Number Theory, 1984–1996*, Amer. Math. Soc., Providence, RI, 1997.
457. A. Matsumoto (Ed.), *Microwave Filters and Circuits*, Academic, NY, 1970.
458. K. Maurin, *Analysis, Vols. I, II*, Reidel, Dordrecht, Holland, 1980.
459. K. Maurin, *General Eigenfunction Expansions and Unitary Representations of Topological Groups*, Polish Scientific Publ., Warsaw, Poland, 1968.
460. F. Mautner, Spherical functions and Hecke operators, *Lie groups and their Representations, Proc. Summer School of the Bolya Janos Math. Soc., Budapest, 1971*, Halsted, N.Y., 1975, 555–576.
461. H. P. McKean, Selberg's trace formula as applied to a compact Riemann surface, *Comm. Pure and Appl. Math., 25* (1972), 225–246; *27* (1974), 134.
462. F. G. Mehler, Über eine mit den Kugel und Cylinderfunktionen verwandte Funktion und ihre Anwendung in der Theorie der Elektricitätsvertheilung, *Math. Ann., 18* (1881), 161–194.
463. J. Mennicke, Vorträge über Selbergs Spurformel, I, U. of Bielefeld, Germany.
464. J. Mennicke, Lectures on discontinuous groups on 3-dimensional hyperbolic space, *Modular Form Conf., Durham, 1983.*
465. A. Messiah, *Quantum Mechanics, Vols. I, II,* Wiley, N.Y., 1961, 1962.
466. C. Meyer, *Die Berechnung der Klassenzahl abelscher Körper über quadratischen Zahlkörpern*, Akademie-Verlag, Berlin, 1957.
467. S. J. Miller and R. Takloo-Bighash, *An Invitation to Modern Number Theory*, Princeton U. Press, Princeton, NJ, 2006.
468. J. Milnor, Hilbert's problem 18, *Proc. Symp. Pure Math., Vol. 28,* American Math. Soc., Providence, RI, 1976, 491–506.
469. J. Milnor, Hyperbolic geometry: the first 150 years, *Bull. American Math. Soc. N.S., 6* (1972), 9–24.
470. S. Minakshisundaram and A. Pleijel, Some properties of the eigenvalues of the Laplacian, *J. Diffffl. Geom., 1* (1967), 43–69.
471. H. Minkowski, *Gesammelte Abhandlungen*, Chelsea, NY, 1911 (reprinted 1967).
472. R. Moeckel, Geodesics on modular surfaces and continued fractions, *Ergodic Theory and Dynamical Systems, 2* (1982), 69–83.
473. S. A. Molchanov, Diffusion processes and Riemannian geometry, *Russ. Math. Surveys, 30* (1975), 1–63.
474. M. I. Monastyrsky and A. M. Perelomov, Coherent states and bounded homogeneous domains. *Reports on Math. Phys., 6* (1974), 1–14.
475. H. L. Montgomery and R. C. Vaughan, *Multiplicative Number Theory I, Classical Theory*, Cambridge U. Press, Cambridge, UK, 2007.
476. L. J. Mordell, On Mr. Ramanujan's empirical expansions of modular functions, *Proc. Cambridge Phil. Soc., 19* (1917), 117–124.
477. C. Moreno, Explicit formulas in the theory of automorphic forms, *Lecture Notes in Math., 626*, Springer-Verlag, NY, 1977, 73–216.
478. C. Moreno and F. Shahidi, The fourth moment of the Ramanujan tau function, *Math. Ann., 266* (1983), 233–239.
479. P. Morse and H. Feshbach, *Methods of Theoretical Physics, Vols. I, II,* McGraw-Hill, NY, 1953.
480. C. J. Mozzochi, *The Fermat Diary*, Amer. Math. Soc., Providence, RI, 2000.
481. C. Müller, *Spherical Harmonics, Lecture Notes in Math., 17,* Springer-Verlag, NY, 1966.
482. D. Mumford, *Curves and Their Jacobians*, U. of Michigan Press, Ann Arbor, MI, 1976.
483. D. Mumford, Geometric Invariant Theory, Springer-Verlag, NY, 1965.
484. D. Mumford, On the equations defining Abelian Varieties, I, II, III, *Inv. Math., 1* (1966), 287–354; *3* (1967), 75–135, 215–244.

485. D. Mumford, *Tata Lectures on Theta, I, II*, Birkhäuser, Boston, 1982, 1984.
486. M. Ram Murty, Oscillations of Fourier coefficients of modular forms, *Math. Ann., 262* (1983), 431–446.
487. M. Ram Murty, On the Sato-Tate conjecture, in N. Koblitz (Ed.), *Number Theory Related to Fermat's Last Theorem*, Birkhäuser, Boston, 1982, 195–205.
488. M. Ram Murty, An introduction to Selberg's trace formula, *J. Indian Math. Soc. (N.S.) 52* (1987), 91{126 (1988).
489. H. Nagoshi, Distribution of Hecke eigenvalues, *Proc. Amer. Math. Soc., 134* (2006), no. 11, 3097–3106.
490. M. A. Naimark, *Linear Differential Operators, Vols. I, II*, Ungar, NY, 1968.
491. W. Narkiewicz, *Algebraic Numbers,* Polish Sci. Pub., Warsaw, 1973.
492. Z. Nehari, *Conformal Mapping*, McGraw-Hill, NY, 1952.
493. E. Nelson, *Dynamical Theories of Brownian Motion,* Princeton U. Press, Princeton, NY, 1967.
494. H. Neuenhöffer, Über die analytische Fortsetzung von Poincaré reihen, *Sitz. Heidelberg Akad. Wiss.* (1973).
495. P. Nicholls, The Ergodic Theory of Discrete Groups, Cambridge U. Press, Cambridge, UK, 1989.
496. M. E. Novodvorsky and I. I. Piatetskii-Shapiro, Rankin-Selberg method in the theory of automorphic forms, *Proc. Symp. Pure Math., 30*, Amer. Math. Soc., Providence, RI, 1976, 297–301.
497. A. Nussbaum, *Applied Group Theory for Chemists, Physicists, and Engineers*, Prentice-Hall, Englewood Cliffs, NJ, 1971.
498. F. Oberhettinger, *Tables of Mellin Transforms,* Springer-Verlag, N.Y., 1974.
499. F. Oberhettinger, *Tables of Bessel Transforms*, Springer-Verlag, NY, 1972.
500. F. Oberhettinger, *Fourier Transforms of Distributions and Their Inverses: A Collection of Tables*, Academic, NY, 1973.
501. F. Oberhettinger, *Fourier Expansions: A Collection of Formulas*, Academic, NY, 1973.
502. F. Oberhettinger and L. Badii, *Tables of Laplace Transforms,* Springer-Verlag, N.Y., 1973.
503. A. M. Odlyzko, Bounds for discriminants and related estimates for class numbers, regulators and zeros of zeta functions: a survey of recent results, Sém. de Théorie des Nombres, Bordeaux, 2 (1990), 119–141.
504. A. M. Odlyzko and H. J. J. te Riele, Disproof of the Mertens conjecture, *J. Reine Angew. Math., 357* (1985), 138–160.
505. A. Ogg, *Modular Forms and Dirichlet Series*, Benjamin, NY, 1969.
506. A. Ogg, Survey of modular functions of one variable, *Lecture Notes in Math., Vol. 320*, Springer-Verlag, NY, 1973, 1–35.
507. A. Ogg, A remark on the Sato-Tate conjecture, *Inv. Math., 9* (1970), 198–200.
508. F. W. J. Olver, *Asymptotics and Special Functions*, Academic, NY, 1974.
509. G. C. Papanicolaou, Wave propagation in a one-dimensional random medium, *SIAM J. Appl. Math., 21* (1971), 13–18.
510. S. J. Patterson, A lattice-point problem in hyperbolic space, *Mathematika, 22* (1975), 81–88; *23* (1976), 227.
511. H. Petersson, Über eine Metrisierung der ganzen Modulformen, *Jahresbericht d. Deutsche Math. Verein., 49* (1939), 49–75.
512. H. Petersson, Über die Entwicklungskoeffizienten der Automorphen Formen, *Acta Math., 58* (1932), 170–215.
513. W. Pfetzer, Die Wirkung der Modulsubstitutionen auf mehrfache Thetareihen zu ungerader Variablenzahl, *Arch. Math., 6* (1953), 448–454.
514. R. S. Phillips and P. Sarnak, On cusp forms for co-finite subgroups of $PSL(2,\mathbb{R})$, *Invent. Math., 80, no. 2* (1985), 339–364.
515. I. I. Piatetskii-Shapiro, Euler subgroups, Lie groups and their representations, in I.M. Gelfand (Ed.), *Summer School of the Bolyai Janos Math. Soc.*, Halsted, NY, 1975, 597–620.

516. I. I. Piatetskii-Shapiro, Cuspidal automorphic representations associated to parabolic subgroups and Ramanujan conjecture, in N. Koblitz (Ed.), *Number Theory related to Fermat's Last Theorem*, Birkhäuser, Boston, MA, 1982.
517. H. Poincaré, *Oeuvres, Vols. I-XI*, Gauthier-Villars, Paris, 1951–53.
518. L. Pontryagin, *Topological Groups*, Moscow, USSR, 1957.
519. C. Popescu, K. Rubin, and A. Silverberg, *Arithmetic of L-Functions, IAS/Park City Math. Series, Vol. 18*, AMS, Providence, RI, 2011.
520. D. L. Powers, *Boundary Value Problems*, 3rd Edition, Saunders (Harcourt Brace), Ft. Worth, TX, 1987.
521. D. M. Pozar, Microwave Engineering, Wiley, New York, NY, 1998.
522. E. Prestini, *The Evolution of Applied Harmonic Analysis*, Birkhäuser, Boston, MA, 2004.
523. G. Purdy, R. and A. Terras, and H. Williams, Graphing L-Functions of Kronecker Symbols in the Real Part of the Critical Strip, *Math. Student, 47* (1979), 101–131.
524. E. T. Quinto et al, *Radon Transforms and Tomography, Cont. Math.*, 278, AMS, Providence, RI, 2001.
525. H. Rademacher, *Topics in Analytic Number Theory,* Springer-Verlag, NY, 1973.
526. H. Rademacher, Eine Bemerkung über die Heckeschen Operatoren $T(n)$, *Abh. Math. Sem. U. Hamburg, 31* (1967), 149–151.
527. H. Rademacher and E. Grosswald, *Dedekind Sums, Carus Math. Monographs*, 19, M.A.A., 1972.
528. J. Radon, Über die Bestimmung von Funktionen durch ihre Integralwerte längs gewisser Mannigfaltigkeiten, *Berichte Sächsische Akad. Wissen. zu Leipzig, 69* (1917), 262–277.
529. K. Ramachandra, Some applications of Kronecker's limit formula, *Ann. of Math, 80* (1964), 104–148.
530. B. Randol, Small eigenvalues of the Laplace operator on compact Riemann surfaces, *Bull. Amer. Math. Soc, 80* (1974), 996–1000.
531. B. Randol, A remark on the multiplicity of the discrete spectrum of congruence subgroups, *Proc. Amer. Math. Soc., 81* (1981), 339–340.
532. R. A. Rankin, *Modular Forms and Functions*, Cambridge University Press, Cambridge, UK, 1977.
533. R. A. Rankin, On horocyclic groups, *Proc. London Math. Soc., (3)4* (1954), 219–234.
534. R. A. Rankin, Contributions to the theory of Ramanujan's function $\tau(n)$ and similar arithmetical functions, I, II, III, *Proc. Cambridge Phil. Soc., 35* (1939), 351–356, 357–372; *36* (1940), 150–151.
535. R. Rankin and J. M. Rushforth, The coefficients of certain integral modular forms, *Proc. Cambridge Phil. Soc., 50* (1954), 305–308.
536. H. Rauch and A. Lebowitz, *Elliptic Functions, Theta Functions, and Riemann Surfaces*, Williams and Wilkins, Baltimore, MD, 1973.
537. J. W. S. Rayleigh, On the character of the complete radiation at a given temperature, *Phil. Mag., 27* (1889) and *Scientific Papers*, Cambridge U. Press, Cambridge, 1902, and Dover, NY, 1964, Vol. 3, 273.
538. M. J. Razar, Values of Dirichlet series at integers in the critical strip, *Lecture Notes in Math., 627*, Springer-Verlag, NY, 1977, 1–10.
539. M. J. Razar, Modular forms for $\Gamma_0(N)$ and Dirichlet series, *Trans. Amer. Math., Soc. 231* (1977), 489–495.
540. M. Reed and B. Simon, *Methods of Modern Mathematical Physics, Vol. I, Functional Analysis*, Academic, NY, 1972.
541. D. St. P. Richards, The central limit theorem on spaces of positive definite matrices, *J. Multivariate Anal. 29* (1989), no. 2, 326–332.
542. B. Riemann, *Collected Works*, Dover, N.Y., 1953.
543. L. Robin, *Functions Sphériques de Legendre et Functions Sphéroidales*, Vol. 3, Gauthier-Villars, Paris, 1959.
544. W. Roelcke, Das Eigenwertproblem der automorphen Formen in der hyperbolischen Ebene I, II, *Math. Ann., 167* (1966), 292–337; *168* (1967), 261–324.

545. W. Roelcke, Über die Wellengleichung bei Grenzkreisgruppen erster Art., *Sitzber. Akad. Heidelberg, Math.-naturwiss. Kl. 1953/55.*
546. W. Roelcke, Über den Laplace-Operator auf Riemannschen Mannigfaltigkeiten mit diskontinuierlichen Gruppen. *Math. Nachr., 21* (1960), 131–149.
547. S. Rosenberg, *The Laplacian on a Riemannian Manifold*, Cambridge U. Press, Cambridge, UK, 1997.
548. Z. Rudnick and P. Sarnak, The behavior of eigenstates of arithmetic hyperbolic manifolds, *CMP, 161* (1994), 195–213.
549. D. Ruelle, Zeta functions and statistical mechanics, *Astérisque, 40* (1976), 167–176.
550. W. Rühl, *The Lorentz Group and Harmonic Analysis*, Benjamin, NY, 1970.
551. W. Rühl, An elementary proof of the Plancherel theorem for the classical groups, Comm. Math. Phys., 11 (1969), 297–302.
552. H. Saito, *Automorphic Forms and Algebraic Extensions of Number Fields, Lectures in Math., 8*, Kinokuniya Book Store, Tokyo, Japan, 1975.
553. Y. Sakamoto, Calculation of Madelung's coefficient of NaCl, *J. Sci. Hiroshima U., A, 16* (1953), 569.
554. P. Samuel, *Algebraic Theory of Numbers,* Houghton-Miffiin, Boston, 1970.
555. P. Sarnak, *Some Applications of Modular Forms*, Cambridge U. Press, Cambridge, UK, 1990.
556. P. Sarnak, Selberg's eigenvalue conjecture: on the occasion of Robert Osserman's retirement, Notices of the Amer. Math. Soc., Vol. 42, No. 11, Nov. 1995, 1272–1279.
557. P. Sarnak, Kloosterman, quadratic forms and modular forms, *Kloosterman Centennial Celebration, Nieuw. Arch. Wiskd., 5/1, nr.4* (Dec., 2000), 385–389.
558. P. Sarnak, The arithmetic and geometry of some hyperbolic three-manifolds, *Acta Math., 151* (1983), 253–295.
559. P. Sarnak, Linking numbers of modular knots, www.math.princeton.edu/sarnak.
560. P. Sarnak, Recent progress on the quantum unique ergodicity conjecture, *Bull. Amer. Math. Soc., 48, No. 2* (April, 2011), 211–228.
561. P. Sarnak and A. Strömbergsson, Minima of Epstein's Zeta Function and Heights of Flat Tori, *Invent. Math., 165* (2006), 115–151.
562. B. Schoeneberg, *Elliptic Modular Functions*, Springer-Verlag, NY, 1974.
563. A. Schuster, On the investigation of hidden periodicities with application to a supposed 26 day period of meteorological phenomena, *Terr. Magn., 3* (1898), 13–41.
564. E. Schwandt, Non-degenerate wavefunctions with wave parameter $\rho = 1$, *J. London Math. Soc., 3* (1971), 183–186.
565. L. Schwartz, *Mathematics for the Physical Sciences*, Hermann, Paris, 1966.
566. L. Schwartz, Théorie des Distributions, Hermann, Paris, 1959.
567. L. Schwartz, Some applications of the theory of distributions, in T.L. Saaty (Ed.), *Lectures on Modern Mathematics,* Wiley, N. Y., 1963, 23–58.
568. R. L. E. Schwarzenberger, *N-dimensional Crystallography,* Pitman, San Francisco, 1980.
569. A. Selberg, Harmonic analysis and discontinuous groups in weakly symmetric Riemannian spaces with applications to Dirichlet series, *J. Indian Math. Soc., 20* (1956), 47–87.
570. A. Selberg, On the estimation of Fourier coefficients of modular forms, *Proc. Symp. Pure Math., 8*, American Math. Soc., Providence, R.I., 1965, 1–15.
571. A. Selberg, Discontinuous groups and harmonic analysis, *Proc. Internatl. Cong. of Math., Stockholm, 1962*, 177–189.
572. A. Selberg, Bemerkungen über eine Dirichletsche Reihe, die mit der Theorie der Modulformen nahe verbunden ist, *Arch. Math. Naturvid., 43* (1940), 47–50.
573. A. Selberg, *Lectures on the Trace Formula*, University of Göttingen, 1954.
574. A. Selberg, *Collected Papers, vols. 1 and 2*, Springer-Verlag, Berlin, 1989 and 1991.
575. M. Senechal, *Quasicrystals and geometry*, Cambridge U. Press, 1995.
576. J.-P. Serre, *A Course in Arithmetic*, Springer-Verlag, NY, 1973.
577. J.-P. Serre, Modular forms of weight one and Galois representations, in A. Fröhlich (Ed.), *Algebraic Number Fields,* Academic, NY, 1977, 193–268.

578. J.-P. Serre, Cohomologie des groupes discrets, *Prospects in Math.*, Princeton U. Press, Princeton, N.J., 1971.
579. J.-P. Serre, Lectures on coefficients of modular forms, Modular Forms Conf., Durham, 1983.
580. J.-P. Serre, *Abelian l-adic Representations and Elliptic Curves*, Benjamin, NY, 1968.
581. J.-P. Serre and H.M. Stark, Modular forms of weight $1/2$, *Lecture Notes in Math., 627*, Springer-Verlag, NY, 1977, pp. 28–67.
582. I.R. Shafarevitch, *Basic Algebraic Geometry*, Springer-Verlag, NY, 1974.
583. A. Shaheen, *Finite Planes and Finite Upper Half Planes: Their Geometry, a Trace Formula, Modular Forms and Eisenstein Series*, Ph.D. Thesis, U.C.S.D., 2005.
584. A. Shaheen, A trace formula for finite upper half planes, *J. Ramanujan Math. Soc., 21* (2006), no. 4, 343–363.
585. T. Shaheen and A. Terras, Fourier expansions of complex-valued Eisenstein series on finite upper half planes, I*nternational J. of Math. and Math. Sciences, Vol. 2006*, Article ID 63918, pp. 1–17.
586. D. Shanks, Five number-theoretic algorithms, in *Proc. 2nd Manitoba Conf. Numerical Math.*, Winnipeg, Manitoba, Canada, 1972, pp. 51–70.
587. C. E. Shannon, Communication in the presence of noise, *Proc. I.R.E., 37* (1949), 10–21.
588. L. A. Shepp and J. B. Kruskal, Computerized tomography: the new medical x-ray technology, *Amer. Math. Monthly, 85* (1978), 420–429.
589. G. Shimura, *Introduction to the Arithmetic Theory of Automorphic Functions,* Princeton U. Press, Princeton N.J., 1971.
590. G. Shimura, On modular forms of half integral weight, *Ann. of Math., 97* (1973), 440–481.
591. G. Shimura, On the holomorphy of certain Dirichlet series, *Proc. London Math. Soc., 31* (1975), 79–98.
592. G. Shimura, Sur les intégrales attachées aux formes automorphes, *J. Math. Soc. Japan, 11* (1979), 291–311.
593. T. Shintani, On evaluation of zeta functions of totally real algebraic number fields at non-positive integers, *J. Fac. Sci. U. Tokyo, 23* (1976), 393–417.
594. T. Shintani, A proof of the classical Kronecker limit formula, *Tokyo J. Math., 3* (1980), 191–199.
595. T. Shintani, On special values of zeta functions of totally real algebraic number fields, *Proc. Internatl. Cong. of Math. Helsinki, 1978*, 591–597.
596. C. L. Siegel, *Topics in Complex Function Theory,* Vols. I,II,III, Wiley-Interscience, NY, 1969–73.
597. C. L. Siegel, *Lectures on Quadratic Forms*, Tata Institute of Fundamental Research, Bombay, 1957.
598. C. L. Siegel, *Analytische Zahlentheorie, Lecture Notes*, U. of Göttingen, 1963–64.
599. C. L. Siegel, *Lectures on Advanced Analytic Number Theory*, Tata Institute of Fundamental Research, Bombay, 1963.
600. C. L. Siegel, *Gesammelte Abhandlungen, Vols. I-IV*, Springer-Verlag, N.Y., 1966, 1979.
601. C. L. Siegel, On the history of the Frankfurt mathematics seminar, *Math. Intelligencer, 1* (1979), 233–246.
602. G. J. Simmons, Cryptology: The mathematics of secure communication, *Math. Intelligencer, 1* (1979), 233–246.
603. I. M. Singer, Eigenvalues of the Laplacian and invariants of manifolds, *Proc. Internatl. Cong. of Math., Vancouver, 1974, Vol. I*, 187–200.
604. I. M. Singer and J. A. Thorpe, *Lectures on Elementary Topology and Geometry*, Scott, Foresman and Co., Glenview, Illinois, 1967.
605. D. Slepian and H. O. Pollak, Prolate spheroidal wave functions, Fourier analysis and uncertainty, I, II, *Bell System Tech. J., 1* (1961), 43–84.
606. N. J. A. Sloane, Binary codes, lattices, and sphere-packings, *Combinatorial Surveys, Proc. 6th British Comb. Conf.,* Academic, London, UK, 1977, 117–164.
607. N. J. A. Sloane, Error-correcting codes and invariant theory: New applications of a 19th century technique, *Math. Monthly, 84* (1977), 82–107.

608. R.A. Smith, The L^2-norm of Maass wave functions, *Proc. Amer. Math. Soc., 82* (1981), 179–182.
609. I. H. Sneddon, *The Use of Integral Transforms,* McGraw-Hill, NY, 1972.
610. A. Sommerfeld, *Partial Differential Equations in Physics*, Academic, NY, 1949.
611. J. Soto-Andrade, Geometrical Gel'fand models, tensor quotients, and Weil representations, *Proc. Symp. Pure Math., 47*, Amer. Math. Soc., Providence, RI, 1987, pp. 305–316.
612. K. Soundararajan, Quantum Unique Ergodicity for $SL_2(\mathbb{Z})\backslash H$, *Ann. of Math. (2)*, (to appear).
613. R. Spira, Calculation of Dirichlet L-functions, *Math. Comp.*, 23 (1969), 489–497.
614. M. Spivak, *Calculus on Manifolds*, Benjamin, NY, 1965.
615. M. Spivak, *A Comprehensive Introduction to Differential Geometry, Vol. II*, Publish or Perish, Berkeley, CA, 1979.
616. F. D. Stacey, *Physics of the Earth*, Wiley, NY, 1977.
617. E. Stade, *Fourier Analysis,* Wiley, Hoboken, NJ, 2005.
618. I. Stakgold, *Boundary Value Problems of Mathematical Physics, Vols. I, II*, MacMillan, NY, 1967.
619. I. Stakgold, *Green's Functions and Boundary Value Problems*, Wiley NY, 1979.
620. J. A. Staniforth, *Microwave Transmission*, The English Universities Press, London, UK, 1972.
621. D. Stanton, An introduction to group representations and orthogonal polynomials, *Orthogonal Polynomials: Theory and Practice,* Edited by P. Nevai and M. E. H. Ismail, Kluwer, Dordrecht, 1990, 419–433.
622. R. J. Stanton and R. A. Tomas, Expansions for spherical functions on noncompact symmetric spaces, *Acta Math., 140* (1978), 251–276.
623. H. M. Stark, Galois theory, algebraic number theory, and zeta functions, in *From Number Theory to Physics*, M. Waldschmidt et al (Eds.), Springer-Verlag, Berlin, 1992, pp. *313–412.*
624. H. M. Stark, The analytic theory of algebraic numbers, *Bull. American Math. Soc., 81* (1975), 961–972.
625. H. M. Stark, On the problem of unique factorization in complex quadratic fields, *Proc. Symp. Pure Math., 7*, American Math. Soc., Providence, RI, 1969, 41–56.
626. H. M. Stark, Values of L-functions at $s = 1$, I-IV, *Advances in Math., 7* (1971), 301–343; *17* (1975), 60–92; *22* (1976), 64–84; *35* (1980), 197–235.
627. H. M. Stark, Class fields and modular forms of weight one, *Lecture Notes in Math., Vol. 601,* Springer-Verlag, NY, 1977, pp. 277–287.
628. H. M. Stark, M.I.T. and U.C.S.D. number theory course lecture notes (unpublished).
629. H. M. Stark, On the zeros of Epstein's zeta function, *Mathematika, 14* (1967), 47–55.
630. H. M. Stark, The Coates-Wiles theorem revisited, in N. Koblitz (Ed.), *Number Theory Related to Fermat's Last Theorem*, Birkhäuser, Boston, 1982, 349–362.
631. H. M. Stark, Fourier coefficients of Maass wave forms, in R.A. Rankin (Ed.), *Modular Forms*, Horwood, Chichester (distrib. Wiley), 1984, 263–269.
632. H. M. Stark, Galois theory, algebraic number theory and zeta functions, in *From Number Theory to Physics*, M. Waldschmidt et al (Eds.), Springer-Verlag, Berlin, 1992, pages 313–393.
633. H. M. Stark and A. A. Terras, Zeta functions of finite graphs and coverings, II, *Adv. in Math., 154* (2000), 132–195.
634. H. M. Stark and D. Zagier, A Property of L -Functions on the Real Line, *J. Number Theory, 12* (1980), 49–52.
635. E. M. Stein and R. Shakarchi, *Fourier Analysis: An Introduction,* Princeton U. Press, Princeton, NJ, 2003.
636. E. Stein and G. Weiss, *Fourier Analysis on Euclidean Spaces*, Princeton U. Press, Princeton, NJ, 1971.
637. W. Stein, *Algebraic Number Theory, A Computational Approach* (http://modular.math.washington.edu/books/ant/ant.pdf).
638. W. Stein, *Modular Forms, a Computational Approach* , Amer. Math. Soc., Providence, RI, 2007.
639. T. J. Stieitjes, Recherches sur les fractions continues, *Ann. Fac. Sci. Toulouse, 8* (1894), 1–22.

640. G. Strang, Wavelets, *American Scientist, 82* (April, 1994), 250–255.
641. G. Strang and G. I. Fix, *An Analysis of the Finite Element Method*, Prentice-Hall, Englewood Cliffs, NJ, 1973.
642. G. Strang and T. Nguyen, *Wavelets and Filter Banks*, Wellesley-Cambridge Press, Wellesley MA, 1996.
643. R. L. Strichartz, *A Guide to Distribution Theory and Fourier Transforms*, CRC Press, Boca Raton, FL, 1994.
644. F. Strömberg, Maass waveforms on $(\Gamma_0(N),\chi)$ (computational aspects), in *Hyperbolic Geometry and Applications in Quantum Chaos and Cosmology* , J. Bolte and F. Steiner, Eds., Cambridge U. Press, Cambridge, UK, 2012, 187–228.
645. A. Strömbergsson, Some remarks on a spectral correspondence for Maass waveforms, *Intl. Math. Res. Notices, 2001, No. 10*, 505–517.
646. A. Strömbergsson, On the uniform equidistribution of long closed horocycles, *Duke Math. J., 123, No. 3* (2004), 507–547.
647. N. Subia, Formule de Selberg et formes d'espaces hyperboliques compactes, *Lecture Notes in Math., 49*7, Springer-Verlag, NY, 674–700.
648. M. Sugiura, *Unitary Representations and Harmonic Analysis*, Wiley, NY, 1975.
649. T. Sunada, Riemannian coverings and isospectral manifolds, *Ann. of Math., 121* (1985), 169–186.
650. P. Swinnerton-Dyer, Conjectures of Birch and Swinnerton-Dyer and of Tate, in T. A. Springer (Ed.), *Proc. Conf. Local Fields*, Springer-Verlag, NY, 1967, 132–157.
651. G. Szegö, Über einige asymptotische Entwicklungen der Legendreschen Funktionen, *Proc. London Math. Soc, 36* (1932), 427–450.
652. K. Takeuchi, Arithmetic triangle groups, *J. Math. Soc. Japan, 29, No. 1* (1977), 91–106.
653. K. Takeuchi, Commensurability classes of arithmetic triangle groups, *J. Fac. Sci. Univ. Tokyo, Sect. 1A, Math., 24* (1977), 201–212.
654. L. A. Takhtadzhyan and A. I. Vinogradov, The Gauss-Hasse hypothesis on real quadratic fields with class number one, *J. Reine Angew. Math., 335* (1982), 40–86.
655. J. D. Talman, *Special Functions,* Benjamin, N.Y., 1968.
656. T. Tamagawa, On Selberg's trace formula, *J. Fac. Sci. U. Tokyo, Sec. I, 8* (1960), 363–386.
657. J. Tate, The work of David Mumford. *Proc. Internatl. Cong. of Math., Vancouver, 1974, Vol. I,* 11–15.
658. J. Tate, The general reciprocity law, *Proc. Symp. Pure Math., 28,* Amer. Math. Soc., Providence, RI, 1976, 311–322.
659. O. Taussky, On a theorem of Latimer and MacDuffee, *Canadian Math. J.,* 1 (1949), 300–302.
660. O. Taussky, On matrix classes corresponding to an ideal and its inverse, *Illinois J. Math., 1* (1957), 108–113.
661. R. Taylor and A. Wiles, Ring-theoretic properties of certain Hecke algebras, *Ann. of Math., 141 (3)* (1995), 553–572.
662. A. Terras, Applications of special functions for the general linear group to number theory, *Sém. Delange-Pisot-Poitou, 1976–1977*, exp 23.
663. A. Terras, Integral formulas and integral tests for series of positive matrices, *Pacific J. of Math., 89* (1980), 471–490.
664. A. Terras, The minima of quadratic forms and the behavior of Epstein and Dedekind zeta functions, *J. Number Theory, 12* (1980), 258–272.
665. A. Terras, Bessel series expansions of the Epstein zeta function and the functional equation, *Trans. American Math. Soc., 183* (1973), 477–486.
666. A. Terras, Some formulas for the Riemann zeta function at odd integer argument resulting from Fourier expansions of the Epstein zeta function, *Acta Arith., 29* (1976), 181–189.
667. A. Terras, *Harmonic Analysis on Symmetric Spaces and Applications, II*, Springer-Verlag, NY, 1989.
668. A. Terras, *Fourier Analysis on Finite Groups and Applications*, Cambridge U. Press, Cambridge, UK, 1999.

669. A. Terras, Comparison of Selberg's trace formula with its discrete analogues, in (ed. Melvyn B. Nathanson), *Proc. DIMACS Workshop on Unusual Applications of Number Theory, DIMACS Series in Discrete Math., Vol. 64*, Amer. Math. Soc., 2004, pp. 213–225.
670. A. Terras, Finite Models for Arithmetical Quantum Chaos, *IAS/Park City Math. Series, Vol. 12*, A.M.S., Providence, RI, 2007, pages 333–375.
671. A. Terras, *Zeta Functions of Graphs*, Cambridge U. Press, Cambridge, UK, 2011.
672. A. Terras and D. Wallace, Selberg's trace formula on k-regular trees and applications, *Internatl. Journal of Math. and Math. Sciences, Vol. 2003, No. 8* (Feb. 2003), 501–526.
673. R. Terras, On the convergence of the continued fraction for the incomplete gamma function and an algorithm for the Riemann zeta function, Preprint.
674. R. Terras, The determination of incomplete gamma functions through analytic integration, *J. Comp. Physics, 31* (1979), 146–151.
675. R. Terras, Generalized exponential operators in the continuation of the confluent hypergeometric functions, *J. Comp. Physics, 44* (1981), 156–166.
676. B. Thaller, *Advanced Visual Quantum Mechanics*, Springer, N.Y., 2005.
677. H. Then, Maass cusp forms for large eigenvalues, *Mathematics of Computation 74* (2005), no. 249, 363–381.
678. E. C. Titchmarsh, *Introduction to the Theory of Fourier Integrals*, Oxford U. Press, Oxford, UK, 1937.
679. E. C. Titchmarsh, *Eigenfunction Expansions Associated with Second Order Differential Equations, Vols. I, II,* Oxford U. Press, Oxford, UK, 1946.
680. E. C. Titchmarsh, *The Theory of Riemann's Zeta Function,* Oxford University Press, Oxford, 1951.
681. P. D. Tiu and D. Wallace, Norm quadratic residue codes, *IEEE Trans. Inform. Theory, 40 (3)*, (1994), 946–949.
682. P. de La Torre and L. J. Goldstein, On the transformation of $\log \eta(\tau)$, *Duke Math. J., 41* (1973), 291–297.
683. J. S. Trefil, *Introduction to the Physics of Fluids and Solids*, Pergamon, NY, 1975.
684. F. Treves, Applications of distributions to PDE theory, *Amer. Math. Monthly, 77* (1970), 241–248.
685. K. Trimèche, Probabilitiés indéfiniment divisibles et théorème de la limite centrale pour une convolution généralisée sur la demi-droite, *C.R. Acad. Sci. Paris, 286* (1978), 63–66.
686. J. Tunnell, Artin's conjecture for representations of octahedral type, *Bull. Amer. Math. Soc., 5* (1981), 173–175.
687. J. Tunnell, On the local Langlands conjecture for $GL(2)$, *Inv. Math., 46* (1978), 179–200.
688. V. N. Tutubalin, On limit theorems for the product of random matrices, *Theory of Probability and its Applications, 15* (1965), 15–27.
689. G. A. Vanasse, *Spectrometric Techniques, I,* Academic, NY, 1977.
690. A. Van Der Poorten, A proof that Euler missed... Apery's proof of the irrationality of $\zeta(3)$, *Math. Intelligencer, 1* (1979), 195–203.
691. B. L. Van Der Waerden, *Group Theory and Quantum Mechanics*, Springer-Verlag, NY, 1974.
692. V. S. Varadarajan, *Lie Groups, Lie Algebras and their Representations*, Prentice-Hall, Englewood Cliffs, NJ, 1974.
693. A. B. Venkov, Spectral theory of automorphic functions, the Selberg zeta function, and some problems of analytic number theory, *Russian Math. Surveys, 34* (1979), 79–153.
694. A. B. Venkov, On the trace formula for $SL(3,\mathbb{Z})$, J. Sov. Math., 12 (1979), 384–424.
695. A. B. Venkov, *Spectral Theory of Automorphic Functions, Proc. Steklov Inst. Math., 153* (1981) Amer. Math. Soc. translation, 1982.
696. A. B. Venkov, V. L. Kalinin, and L. D. Faddeev, A nonarithmetic derivation of the Selberg trace formula, *J. Soviet Math., 8* (1977), 171–199.
697. M. Vergne, Representations of Lie groups and the orbit method, in B. Srinivasan and J. Sally (Eds.), *Emmy Noether in Bryn Mawr,* Springer-Verlag, NY, 1983, 59–101.
698. H. Verrill, Subgroups of $PSL_2(\mathbb{R})$, Chapter in The Magma Handbook, Vol. 2, J. Cannon, W. Bosma Eds., (2001), 233–254.

699. M.-F. Vignéras, Séries thèta des formes quadratiques indéfinies, *Lecture Notes in Math., 627*, Springer-Verlag, NY, 1977, 227–240.
700. M.-F. Vignéras, L'équation fonctionelle de la fonction zêta de Selberg de la groupe modulaire $PSL(2,\mathbb{Z})$, *Astérisque, 61* (1979), 235–249.
701. M.-F. Vignéras, Variétés riemanniennes isospectrales et non isométriques, Ann. of Math., (2) 112 (1980), 21–32.
702. M.-F. Vignéras, *Arithmétique des Algèbres de Quaternions, Lecture Notes in Math., 800,* Springer-Verlag, NY, 1980.
703. M.-F. Vignéras, Quelques remarques sur la conjecture $\lambda_1 \geq \frac{1}{4}$, *Sém Théorie des Nombres, Paris 1981–82,* M.-J. Bertin (Ed.), Birkhäuser, Boston, 1983, 321–343.
704. N. J. Vilenkin, *Special Functions and the Theory of Group Representations, Translations of Math. Monographs, 22,* American Math. Soc., Providence, RI, 1968.
705. A. D. Virtser, Central limit theorem for semisimple Lie groups, *Theory of Probability and its Applics., 15* (1970), 667–687.
706. V. S. Vladimirov, *Equations of Mathematical Physics*, Dekker, NY, 1971.
707. D. Vogan, *Unitary Representations of Reductive Lie Groups, Annals of Mathematics Studies*, Princeton University Press, Princeton, New Jersey, 1987.
708. J. Voight, Computing fundamental domains for Fuchsian groups, *J. de Theorie des Nombres de Bordeaux, 21* (2009), 467–489.
709. S.M. Voronin, On analytic continuation of certain Dirichlet series, *Proc. Steklov Inst. Math., 157* (1983), 25–30.
710. A. Voros, Spectral zeta functions, in *Zeta Functions in Geometry (Proceedings, Tokyo, Japan, August 1990), Advanced Studies in Pure Mathematics, 21* (1992) 327–358.
711. D. I. Wallace, *Selberg's trace formula and units in higher degree number fields*, Ph.D. Thesis, U.C.S.D., 1982.
712. D. I. Wallace, Conjugacy classes of hyperbolic matrices in $SL(n,\mathbb{Z})$ and ideal classes in an order, *Trans. Amer. Math. Soc., 283* (1984), 177–184.
713. D. I. Wallace and J. BelBruno, *The Bell That Rings Light: A Primer in Quantum Mechanics And Chemical Bonding*, World Scientific, Singapore, 2006
714. N. Wallach, Lecture Notes, 1981 NSF-CBMS Regional Conf. on Representations of Semi-Simple Lie Groups and Applics. to Analysis, Geometry and Number Theory (unpublished).
715. N. Wallach, *Harmonic Analysis on Homogeneous Spaces*, Dekker, NY, 1973.
716. A. Wangerin, Theorie der Kugelfunktionen und der verwandten Funktionen, insbesondere der Laméschen und Besselschen (Theorie spezieller durch lineare Differential-gleichungen definierter Funktionen), *Enzykl. der Mathematischen Wissenschaften, Vol. 2, Pt. 1, No. 2,* Teubner, Leipzig, 1904–1916, 699–762.
717. G. Warner, *Harmonic Analysis on Semi-Simple Lie Groups, Vols. I, II*, Springer-Verlag, NY, 1972.
718. G. Warner, Selberg's trace formula for non-uniform lattices: the $\mathbb{R}$-rank one case, *Studies in Algebra and Number Theory, Advances in Math., Supp. Studies, 6* (1979), 1–142.
719. M. Watkins, Class number of imaginary quadratic fields, *Math. of Computation, 73* (2003), 907–938.
720. G. Watson, *Statistics on Spheres,* Wiley, NY, 1983.
721. G. N. Watson, *A Treatise on the Theory of Bessel Functions,* Cambridge U. Press, London, UK, 1962.
722. A. Wawrzyńczyk, *Group Representations and Special Functions*, Polish Sci. Pub., Warsaw, Poland, and Reidel, Dordrecht, Holland, 1984.
723. N. Wax (Ed.), *Selected Papers on Noise and Stochastic Processes*, Dover, NY, 1954.
724. H. Weber, *Lehrbuch der Algebra, Vols. I-III*, Chelsea, N.Y., 1961.
725. A. Weil, *L'intégration dans les Groupes Topologiques*, Hermann, Paris, France, 1965.
726. A. Weil, *Basic Number Theory,* Springer-Verlag, N.Y., 1974.
727. A. Weil, *Elliptic Functions According to Eisenstein and Kronecker*, Springer-Verlag, NY, 1976.

728. A. Weil, *Dirichlet Series and Automorphic Forms, Lecture Notes in Math., 189*, Springer-Verlag, NY, 1971.
729. A. Weil, *Oeuvres Scientifiques, Collected Papers, 1926–1978, Vols. I-III*, Springer-Verlag, NY, 1979.
730. H. Weyl, *Gesammelte Abhandlungen*, Springer-Verlag, NY, 1968.
731. H. Weyl, *The Theory of Groups and Quantum Mechanics*, Dover, NY, 1928.
732. H. Weyl, *The Classical Groups, Their Invariants and Representations*, Princeton U. Press, Princeton, NJ, 1946.
733. H. E. White, *Introduction to Atomic Spectra*, McGraw-Hill, NY, 1934.
734. E. T. Whittaker and G. N. Watson, *A Course of Modern Analysis*, Cambridge U. Press, Cambridge, UK, 1927.
735. D. V. Widder, *The Laplace Transform,* Princeton U. Press, Princeton, NJ, 1941.
736. N. Wiener, *The Fourier Integral and Certain of its Applications*, Cambridge U. Press, London, UK, 1933.
737. N. Wiener, *Selected Papers*, M.I.T. Press, Cambridge, MA, 1964.
738. E. P. Wigner, *Group Theory and its Application to the Quantum Mechanics of Atomic Spectra*, Academic, NY, 1959.
739. E. P. Wigner, Random matrices in physics, *S.I.A.M. Review, 9, No. 1* (1967), 1–23.
740. A. Wiles, Modular elliptic curves and Fermat's last theorem, *Ann. of Math., 141 (3)* (1995), 443–551.
741. F. Williams, *Topics in Quantum Mechanics*, Birkhäuser, Boston, 2003.
742. H. Williams and J. Broere, A computational technique for evaluating $L(1,\chi)$ and the class number of a real quadratic field, *Math. Comp., 30* (1976), 887–893.
743. K. Wohlfahrt, Über Operatoren Heckescher Art bei Modulformen reeller Dimension, *Math. Nachr., 16* (1957), 233–256.
744. S. A. Wolpert, The eigenvalue spectrum as moduli for compact Riemann surfaces, Bull. Amer. Math. Soc., 83 (1977), 1306–1308.
745. S. A. Wolpert, The length spectra as moduli for compact Riemann surfaces, Ann. of Math., 109 (1979), 323–351.
746. B. G. Wybourne, Classical Groups for Physicists, Wiley, NY, 1974.
747. T. Yamazaki, On Siegel modular forms of degree 2, *Amer. J. of Math., 98* (1973), 39–53.
748. K. Yosida, *Functional Analysis,* Springer-Verlag, NY, 1968.
749. D. Zagier, Elliptic Modular Forms and Their Applications, in *The 1-2-3 of Modular Forms: Lectures at a Summer School in Nordfjordeid, Norway*, Edited by J. H. Bruinier, G. van der Geer, G. Harder, D. Zagier, Springer, Berlin, 2008, pp. 1–103.
750. D. Zagier, Eisenstein series and the Riemann zeta function, Eisenstein series and the Selberg trace formula I, in *Automorphic Forms, Representation Theory and Arithmetic, Bombay Colloquium, 1979*, Springer-Verlag, NY, 1981, 275–302, 303–355.
751. D. Zagier, A Kronecker limit formula for real quadratic fields, *Math. Ann., 231* (1975), 153–184.
752. D. Zagier, Modular forms whose coefficients involve zeta functions of quadratic fields, *Lecture Notes in Math., 627*, Springer-Verlag, NY, 1977, 105–169.
753. D. Zagier, Correction to "The Eichler-Selberg trace formula on $SL(2,\mathbb{Z})$," *Lecture Notes in Math, 627,* Springer-Verlag, N.Y., 1977, 171–173.
754. J. M. Ziman, *Principles of the Theory of Solids,* Cambridge U. Press, Cambridge, UK, 1972.
755. H. Zirin, *The Solar Atmosphere,* Blaisdell, Waltham, MA, 1966.
756. I. J. Zucker, Madelung constants and lattice sums for invariant cubic lattice complexes and certain tetragonal structures, *J. Phys. A:Math. Gen., 8* (1975), 1734–1745.
757. A. Zygmund, *Trigonometric Series*, Warsaw, Poland, 1935.

Index

A
Abel's integral equation, 24
abelian extension, *see* class field or abelian extension
abelian integral, 248
abelian theorem, 23
abelian variety, 247, 300
addition formula, 131, 132, 135
adelic theory, 131, 227, 248, 284, 286, 300, 366
adjacency operator, 92, 223
adjoint, 66, 138, 322
algebraic integer, 71, 202
algebraic number field, 71, 202, 217, 256, 265, 273, 298, 304, 317, 360, 366
algebraic variety, 301, 306
algebriac number field, 76
ALGOL program, 67
aliasing, 49
almost periodic function, 264
analytic continuation, 60, 63, 64, 237, 262, 267, 272, 274, 293, 298, 313, 316, 340, 365
analytic function, 199
analytic functionals, 178
arc length, 109–111, 142, 150–152, 156, 248, 270
area element on symmetric space, 111, 153, 205, 209
area fundamental domain modular group, 205
area unit sphere in euclidean space, 5
Artin L-function, 75, 300, 317, 366
Artin reciprocity law, 83, 265
Arzelà-Ascoli theorem, 331
associated Legendre function, P_s^a, 113, 114, 172, 173
asymptotics of eigenvalues of the Laplacian, 45, 360
asymptotics of special functions, 135, 167, 170, 172, 175, 192, 321, 336
asymptotics/functional equations principle, 168, 169, 175
Atiyah-Singer index theorem, 366
atomic physics, 118, 123, 165, 189, 203
autocovariance or autocorrelation, 50
automorphic form, 200, 227, 228, 238, 240, 259, 262
automorphic function, 199, 202, 228, 236, 241

B
Balmer series of spectral lines, 119–121
band limited, 36, 55
Barnes double gamma function, 365
base change, xi, 300
$\mathbb{Z}$-basis, *see* integral basis
basis problem for modular forms, 234, 254, 304
Bernoulli number B_n, 82, 257, 266
Bessel function
  H_ν, 140
  I_s, 142, 238, 341
  J_ν, 134, 135, 140, 191, 194
  K_s, 17, 44, 62, 67, 142, 165, 167, 168, 170, 260, 262, 277, 281, 287, 289, 290, 295, 337
beta function, 173
binary code, 252
binary quadratic form, 79, 80, 153, 202, 267, 366
Bochner-Hecke formula, 135
body-centered cubic lattice, 83
Borel sets, 25
Borel-Weil-Bott theorem, 129
boson pair creation, 203

A. Terras, *Harmonic Analysis on Symmetric Spaces—Euclidean Space, the Sphere, and the Poincaré Upper Half-Plane*, DOI 10.1007/978-1-4614-7972-7,
© Springer Science+Business Media New York 2013

bound states, 119
boundary of a symmetric space, 178
bounded variation, 23
box dimension, 105
branched Riemann surface, 214
Brauer-Siegel theorem, 74, 81
Brownian motion, 29, 318
Bruhat decomposition, 264

C
calculus of variations, 109
CAT scanner, 144
Cauchy principal value, 142, 143
Cauchy problem, 20
Cauchy residue theorem, 23, 40, 76, 238, 266, 321
Cauchy-Riemann equations, 151
Cayley transform, 156, 157, 198, 205
central element, 344, 352
central limit theorem, 28, 29, 118, 195
centralizer, 254, 345, 346, 354
Cesàro sum, 3, 34
chaos, 316, 365
character, 92
character of a representation, 84, 128, 129, 224, 225, 254, 289, 297, 301
characteristic function, 25, 26, 190, 192, 193
circle method, 239
circle problem, 45, 334
class field or abelian extension, 73, 83, 202, 231, 241
class group or ideal class group, 73, 77
class number, 73, 74, 80–82, 207, 208, 239, 265, 345, 347, 363
class-one representation, 130
Clebsch-Gordon series, 128
closed or periodic geodesic, 334, 346, 350, 364
coding theory, 252
coherent states, 203
cohomology, 129, 315
combinatorial Laplacian, 92
compact fundamental domain, 197, 214, 221, 254, 319, 351
compact group, 108, 127, 128, 130, 217, 330
compact operator, 53, 331
compact symmetric space, 108
compactification, 178, 214
completely reducible representation, 126
computerized tomography, 133, 147
confluent hypergeometric function, 62, 69
conformal mapping, 151, 202, 236, 240, 248, 249
congruence subgroup, 200, 211, 236, 244, 286, 299, 306, 310, 316, 332
congruence subgroup problem, 211
conical function, 173
conjecture
  Artin, 300, 317, 366
  Cartier, 287
  Fermat, 70, 150
  Gauss, 267, 365
  Lehmer, 237
  Mertens, 305
  moonshine, 242
  Pólya-Hilbert, 284
  QUE, quantum unique ergodicity, 281
  Ramanujan, 93, 236, 255, 305, 306
  Ramanujan-Petersson, 236, 276, 278, 284–286, 313, 334
  Roelcke-Selberg, 288, 326
  Sato-Tate, 237, 251
  Selberg eigenvalue, 286
  Serre-Sato-Tate, 237
  Stark, 74
  Taniyama, Shimura, Weil, 300
  Weil, 236
conjugacy class, 83, 344–346, 352, 355, 362, 364, 367, 376
conjugate, 71, 72, 76–80, 82, 83, 218, 222, 292
constant term in Fourier expansion of Maass wave form, 264, 268, 269, 288, 322, 323
continued fraction, 68, 73, 205, 315, 349
continuous spectrum, 119, 138, 258, 283, 286, 320, 326, 327, 340, 351
convergence of sequence of distributions, 2, 6, 42
convolution, 3, 4, 6–8, 10, 12, 14, 16, 22, 24, 26, 29, 43, 44, 49, 89, 131, 133, 181, 186, 190, 224, 326–329, 331, 343
correspondence
  between classes of Riemann surfaces and points in fundamental domain, 209
  between ideal classes and points in fundamental domain, 207
  between ideal classes and quadratic forms, 202
  between modular forms and Dirichlet series, 292, 294, 295, 298–300
  between spaces of automorphic forms, 248
  Jacquet-Langlands, 288
Courant minimax principle, 283
covering transformation, 200
cryptography, 70, 223, 251, 252
crystallography, 48, 83, 86
CsCl, 83

cubic equation, 235, 250
cubic lattice, 83, 253
cubic number field, 75
curvature, 153, 196, 280
cusp form, 231, 235, 238, 245, 247, 254, 255, 257, 259, 268, 269, 272, 273, 275–278, 280, 284–288, 291, 296, 297, 300, 303–307, 313, 315, 320, 326, 329, 331, 334, 341, 343, 361, 366
cusp of fundamental domain, 213, 260, 278, 316
cyclic group, 73, 74, 83, 213, 346
cyclotomic field, 70, 73, 75, 83

D
$\mathscr{D}$, test functions, 2
$\tilde{\mathscr{D}}$, test functions, 41
d'Alembert solution wave equation, 14
Dedekind eta function, 230, 238, 250, 264, 374
Dedekind sum, 238, 374
Dedekind zeta function, 75, 76, 80, 81, 273–275, 300, 317, 366
degeneracy, 120
degree of extension of number fields, 71, 75, 77, 305, 341
degree of graph, 91, 93
degree of representation, 125, 130, 242, 287
degree of spherical harmonic, 114
densely wound line in torus, 101, 102, 348
densest lattice packing of spheres, 253
density function of a random variable, 25–28, 190, 191, 193, 195
density plot, 37, 97, 116, 230, 273, 280
derivative of a distribution, 4
determinant one surface in positive matrix space, $\mathscr{S}\mathscr{P}_n$, 153, 154, 260
diamond, 83
different, 80
differentiable manifold, 108
dihedral group, 83
dimension of a fractal, 105
dimension space of automorphic or modular forms
  holomorphic, 233, 235, 305
  Maass cusp forms, 287, 288
δ, Dirac delta distribution, 1, 2, 4–6, 8, 16–19, 22, 42, 49, 57, 99, 101, 133, 139, 141, 142, 169, 175, 188, 318
$\tilde{\delta}$, Dirac delta distribution, 42
direct sum of representations, 126
Dirichlet *L*-function, 274, 297
Dirichlet kernel, 4, 33
Dirichlet polygon, 209
Dirichlet problem, 36, 56, 58, 184, 200, 273, 275, 288, 297
Dirichlet series, 60, 65, 86, 293–300, 302, 304, 305, 307, 313, 315, 325, 340
Dirichlet unit theorem, 74, 79, 81
discontinuous subgroup, 147, 150, 199, 210
discrete or point spectrum, 119
discrete spectrum, 119, 138, 169, 258, 320, 326, 340, 361, 362, 365
discrete subgroup, 81, 197, 200, 209, 211, 214, 260
discriminant
  modular form, *see* Δ, discriminant modular form
  of number field, 72–76, 207, 208, 235, 238, 240, 315, 367
  of quaternion algebra, 221
Δ, discriminant modular form, 230, 231, 234–236, 238–240, 247, 268, 299, 305, 315
dispersion of a random variable, 192
distribution
  charge, 249
  generalized function, 2, 9, 15, 41, 43, 98, 99, 101, 129, 188, 318
    surface layer, 19
  heat, 14, 15, 133, 185, 342
  joint of 2 random variables, 25
  normal or Gaussian, 27–29, 133, 188, 192, 193, 195
  of eigenvalues of Δ, 288
  of horocycles, 312, 314
  of random variable, 25, 190, 280
  of solutions of quadratic congruences, 341
  potential, 249
  semi-circle, Sato-Tate, 237
  uniform, 36, 101, 278, 341
divisor function, 230, 255, 263, 264
doubly periodic function, 200, 235, 248, 250
drum, 36, 317, 318, 365
dual, 80, 129, 130, 143, 178, 334

E
earthquake, 107, 118
Eichler-Selberg trace formula, 306, 317, 366
eigenfunctions, 20, 31, 36, 45, 55–59, 62, 92, 114, 120, 121, 126, 128, 131, 136, 141, 164, 165, 172, 178, 202, 224, 258, 260, 272, 273, 275, 276, 280, 281, 283, 284, 297, 305, 310, 312, 317, 320

eigenvalues, 23, 36, 39, 45, 54, 56, 58, 59, 88, 89, 92, 93, 96, 112, 114, 115, 119, 120, 126, 131, 138, 142, 155, 224, 237, 268, 273, 275, 276, 278, 280, 281, 283, 284, 286, 287, 317, 320, 329, 334, 346, 347, 362, 365
E_s, nonholomorphic Eisenstein series, 260–263, 267, 269, 271, 284, 312–314, 320, 321, 323, 325, 333, 336, 340–342, 344, 356
Eisenstein series, 66, 82, 229, 230, 234, 236, 240, 253–255, 259–264, 267–269, 271, 272, 285, 288, 295, 298, 304, 312, 313, 316, 320, 321, 323, 325, 339
electromagnetic spectrum, 48, 156
elementary divisor theory, 72, 79
elementary row and column operations, 79
elliptic
  curve, 209, 235, 237, 242, 250–252, 265, 267, 284, 300, 301
  element, 214, 317, 345, 354, 361
  fractional linear transformation, 204, 213
  function, 235, 247–250, 257, 258
  geometry, 112, 147, 149, 153
  integral, 200, 235, 247, 248
  partial differential equation, 259
  point, 213, 214, 240
elliptic fractional linear transformation, 213
energy level, 48, 58, 89, 119
Epstein zeta function, 33, 44, 58, 63, 64, 67, 69, 73, 76, 79, 82, 86–88, 229, 259, 260, 262, 273, 283, 292, 325, 339, 341
equation of degree 5, 202
equivalent
  geodesics, 204
  matrices, 163, 309
  points, 203, 213, 260, 326
  positive definite matrices, 243
  representation, 126–128
  Riemann surfaces, 209
ergodic theory, 36, 104, 280, 349, 350
Escher, 198
η, Dedekind eta function, 230, 238, 264, 374
Euclid's fifth postulate, 112, 152
Euclid's second postulate, 112
Euclidean algorithm, 204
Euclidean group, 83, 90, 136, 148, 197, 367
Euler
  –Maclaurin summation formula, 362
  -Lagrange equation, 110
  angle decomposition, 116
  constant, 264, 357
  formula for $\zeta(2n)$, 82, 265
  formula for $\Gamma(s)$, 5, 292
  formula for genus, 214
  function, 75
  Lagrange equation, 110
  product, 63, 75, 76, 302, 304, 305, 307, 311, 362
even integral positive matrix, 243, 256
Ewald's method of theta functions, 87
excited state, 120
expectation or mean, 26, 191, 192

F
Farey fractions, 315
fast Fourier transform or FFT, 31, 46, 90, 94
Féjer kernel, 3, 34
Feynman integrals, 319
Fibonacci tiling, 98
finite analogue of Euclidean distance, 90
finite Dirichlet polygon, 225
finite Eisenstein series, 291
finite element method, 54
finite Euclidean graph, 91
finite Euclidean space, 90
finite fundamental domain, 289
finite general linear group, 222, 292, 374
finite geodesic, 222
finite horocycle, 222
finite non-Euclidean distance, 222
finite rotation group, 222
finite simple group, 242
finite symmetric space, 90
finite tessellation, 225
finite trace formula, 374, 376
finite upper half-plane, 221, 223, 227, 289, 374
finite upper half-plane graph, 223
Fischer–Griess monster group, 242
Fourier analysis on symmetric space, 9, 117, 178
Fourier analysis on the fundamental domain, 31, 39, 333, 352
Fourier coefficient, 31, 40
Fourier inversion, *see* inversion of a transform
Fourier series, 3, 4, 6, 13, 23, 31–37, 39–41, 43, 44, 86, 94, 101, 118, 130, 176, 180, 196, 228, 239, 257, 258, 262, 269, 320, 325
  generalized, 54
  non-Euclidean, 316, 342
Fourier transform, 9–12, 14–19, 21, 23, 25, 26, 28, 29, 31, 36, 40, 42, 43, 46, 49–51, 59, 61, 65, 96, 98, 99, 101, 111, 133–135, 137, 143, 144, 148, 176, 190, 196, 332, 338
  on $\mathbb{Z}/n\mathbb{Z}$, 94

Fourier-Bessel series, 141
fractal, 104–106
fractional ideal, 73, 80
fractional linear transformation, 68, 150–153, 156, 203, 204, 210, 212, 213, 222, 327, 343, 344
Frobenius reciprocity law, 128
Fubini theorem, 12
Fuchsian group, 210, 211, 217, 254, 319
functional equation, 64, 66, 80, 88, 135, 140, 141, 167, 169, 172, 175, 179, 262, 274, 293–296, 298–300, 321, 324, 325, 360, 362, 365
fundamental domain, 31, 147, 197–199, 202–205, 207–216, 218–221, 225, 227, 228, 231–233, 240, 241, 254, 260, 265, 268–273, 275, 278, 280, 286, 310, 312–314, 316, 319, 320, 326, 333, 337, 339, 343, 346, 348–351, 353, 365
fundamental function or Hauptmodul, 240
fundamental or Poincaré group, 108
fundamental solution
  heat equation, 3, 18, 29, 133, 186–191, 193, 195, 342, 361
  Laplace equation or Green's function, 5
  wave equation, 19
  Schrödinger wave equation, 319
fundamental unit, 73–76, 78, 79, 81, 82, 334, 346, 347, 363, 364
Funk-Hecke theorem, 132, 133

G
$\Gamma(N)$, 211
$\Gamma_0(N)$, 200, 244
Gauss
  -Bonnet formula, 112, 153, 265
  distribution, 133, 192, 237
  hypergeometric function, 173–175, 339
  kernel, 3, 10, 14, 27–29, 189
  sum, 267, 289, 290, 299
Gelfand criterion, 89
Gelfand pair, 89
general linear group, 62, 74, 76, 79, 83, 125, 130, 196, 223–225, 227, 243, 244, 286, 289, 291, 300, 334, 375
generators of groups, 204, 211, 220, 299
generators of quaternion algebra, 219
genus, 209, 214, 215, 221, 247, 334
geodesic, 111, 112, 149, 151–153, 156, 208, 209, 215, 254, 280, 315–317, 334, 346–349, 364, 366, 367
  -reversing isometry, 89, 154
  flow, 280, 350, 367
  polar coordinates, 154, 155, 171, 173, 174, 187, 354
  polygon, 204
Gibbs phenomenon, 34, 49, 118
Girard formula, 112
golden ratio, 97
Green's
  function or resolvent kernel, 5, 56–58, 60, 61, 118, 131, 138–142, 254, 327, 333, 337–342
  theorem, 5, 115, 167, 187, 268, 270, 272
Grenzkreisgruppe, 210
grossencharacter, 82, 268, 298, 300
group representation, 20, 48, 92, 125–131, 163, 203, 224, 225, 227, 276, 286, 300, 317, 366, 376

H
Haar measure, 66, 127, 128, 130, 132, 133, 181, 327, 328, 330
Hamming distance, 252
Hankel
  inversion formula, 137, 141, 169
Hankel function, 140
Harish or Abel or horocycle transform, 182, 352, 357, 361, 363
Harish transform, 352
harmonic analysis, *see* Fourier analysis
harmonic function, 170, 270
harmonic polynomial, 114
Hasse-Weil zeta function, 301
Hauptkreisgruppe, 210
Hauptmodul or fundamental function, 240
Hausdorff dimension, 106
heat equation, 3, 13, 18, 23, 29, 133, 164, 186–191, 193, 195, 317, 342, 361
heat kernel, 164, 165, 186, 193
Heaviside step function, 4, 18
Hecke
  algebra, 225
  Bochner-Hecke formula, *see* Bochner-Hecke formula
  correspondence, 292, 294, 299
  Funk-Hecke theorem, *see* Funk-Hecke theorem
  integral formula, 77, 79, 81, 82
  L-function, 82, 268, 284, 298, 300, 360
  operator, 276, 278, 284, 285, 294, 303–306, 308, 312, 314, 315, 317, 366
  points, 278
  triangle group, 216
Heisenberg uncertainty principle, 21

Helgason transform, 178, 180, 182, 183, 187, 328, 329, 333, 335, 344, 353, 363
Hermite polynomials, 59
hidden periodicities, 48, 50
highest point method, 203
highest weight, 128
Hilbert transform, 144
Hilbert's 12th and 18th problems, xi, 70, 74, 82, 197, 202
Hilbert–Schmidt operator, 53
Hilbert-Schmidt operator, 53, 54, 57, 58
Højendal method for Madelung constant, 86
homogeneous space, 108, 153, 154
homology, 315
horocycle, 181, 182, 280, 312, 314, 316, 348, 349, 352
Huyghen's principle, 20
hydrogen, 118, 120, 121, 123, 196, 320
hyperbolic
  3-space, 189
  element of $SL(2,\mathbb{R})$, 254, 255, 341, 345–348, 353, 362, 364
  fractional linear transformation, 213
  geometry, 149, 152, 153, 156, 208
  group, 215
  triangle, 153
  upper half-plane, H, 150
hyperbolic fractional linear transformation, 213
hyperfunction, 178
hypergeometric function, *see* Gauss or confluent hypergeometric function

I

I_s, *see* Bessel function
ideal, 72, 73, 75–77, 79–81, 202, 205, 207, 345
ideal class group, 73, 77, 208, 241
images, method of, 44, 57, 58, 60, 337, 342
impedance, 157, 160
incomplete gamma function, 62, 64, 65, 67–69, 81, 88, 293, 296, 298
incomplete theta series, 321–324
independent random variables, 7, 25–29, 190, 192, 193, 195, 237
indicator function of a set, 195, 314
induced representation, 128, 224, 375
Infeld-Hull factorization method, 58
instrument function, 49
integral basis, 72
integral test, 33, 64, 340
interferogram function, 48
intertwining operator, 126
invariant differential operator, 66, 111, 153, 178
invariant integral, 127, 128, 328
invariant random variable under rotation, 189–192, 195
inversion in a sphere, 57, 249
inversion of a transform
  Fourier, 10–12, 14, 16, 18, 20, 26, 36, 40, 41, 49, 137, 180
  Fourier on Lie group, 129, 196
  Fourier on symmetric space, 175, 354
  Hankel, 137
  Helgason, 178, 180, 183, 353
  Kontorovich-Lebedev, 142, 168, 175, 176
  Laplace, 22, 23
  Mehler-Fock, 175–177, 181
  Mellin, 61, 62, 169, 183, 293, 294, 297, 321, 324
  Radon, 142, 144
irreducible representation, 126, 128, 131, 287
isometric circle, 210
isometric Riemann surface, 366
Iwasawa decomposition, 331

J

J_ν, *see* Bessel function
Jacobi derivative formula, 247
Jacobi identity for Δ, 238
Jacobi theta function, 246, 253
Jacobi transformation formula for theta function, 44
Jacobi triple product formula, 250
Jacobi-Abel functions, 247
Jacobian elliptic function, 248, 249
Jordan form, 212, 344, 345

K

K, *see* maximal compact subgroup
K_s, *see* Bessel function
K bi-invariant function, 224, 328
K-invariant function, 181–183, 189, 190, 192–195, 343, 344, 352, 353
K-theory, 82
kernel of integral operator, 3, 4, 33, 34, 41, 53, 54, 56, 57, 169, 175, 288, 327, 337
Kirchoff's formula for wave equation, 20
Kleinian group, 210
Kloosterman sum, 93, 254, 290, 313
Kodaira–Titchmarsh formula (Stieltjes-Stone also), 137
Kodaira-Titchmarsh formula (Stieltjes-Stone also), 138, 140, 327

Kontorovich-Lebedev transform, 142, 164, 168–171, 175–177
Korteweg–DeVries equation, 257
Korteweg-DeVries equation, 258
Kronecker limit formula, 264, 265
Kronecker symbol, 273
Kronecker theorem, 101

L

L-function, 72, 74, 75, 81–83, 227, 237, 265, 268, 273, 274, 283, 284, 299–301, 310, 317, 360, 366
$\mathit{\Delta}$, Laplace operator, 110–112, 114, 115, 117–119, 133, 136
Δ, Laplace operator, 5, 18, 31, 36–38, 45, 56–58, 60, 92, 155, 164, 167, 170, 173, 184, 185, 187, 258–260, 263, 267, 270, 272, 273, 280, 283, 286, 288, 297, 317, 318, 320, 321, 323, 326, 328, 329, 331, 333, 337, 338, 340–342, 347, 360, 362, 365, 366
Δ, Laplace operator, 153
Laplace series of spherical harmonics, 118
lattice space, 367
least squares, method of, 51
Lebesgue
  dominated convergence theorem, 6, 11, 12, 188, 195
  integrable function, 2, 15, 143, 349
  integrable functions, $L^1(X)$, 7, 11, 12, 32, 127
  integral, 11, 13
  measurable function, 139
  measure, 2, 25, 101, 322
  square integrable functions, $L^2(X)$, 11
left G-action, 154
Legendre function
  P_s, 113, 114, 172–177, 191, 192, 224, 329, 338
  Q_s, 338, 339
Legendre symbol, 244
lemniscate integral, 200
length spectrum of Riemannian manifold, 346, 347, 364, 366
length standard, 46
level of congruence subgroup, 211, 215, 244, 247, 286, 301, 306
Lie group, 89, 123, 125–129, 150, 168, 189, 196, 199, 203, 242, 244, 250, 253, 322, 356
limit point of a discrete group, 210
limit-point and limit-circle case in ODE, 140
linking number of knots, 371
Liouville theorem
  in complex analysis, 241
  on conformal maps in 3-space, 249
Lobatchevsky upper half-plane, *see* Poincare upper half plane, *see* Poincaré upper half-plane, H
locally Euclidean space, 89, 125
locally integrable function, 2, 4, 6–8
Lorentz knot, 368
Lorentz-type group, x, 123, 150, 162, 163
Lyman series of spectral lines, 120

M

Maass cusp form, 268, 272, 273, 276, 280, 284, 287, 288, 307
Maass waveform, 259, 262, 269, 276, 278, 280, 284, 295–297, 306, 314, 361
Maass-Selberg relations, 272
Madelung constant, 86–88
magnetic field, 120, 121, 123, 147
matched load in an electrical network, 157, 158
matrices associated wth electrical circuits, 160
matrix entry of a group representation, 126, 131
maximal compact subgroup, K, 153, 154, 178, 181, 187, 190, 195, 343, 344, 352, 353
maximum principle for harmonic functions, 270
Maxwell’s equations, 123, 150
mean, 26, 28, 50, 189, 191, 192, 195
Mehler-Fock transform, 164, 169, 175–177, 181, 355
Mellin transform, M, 61, 62, 77, 169, 176, 182, 183, 200, 244, 263, 292–297, 299, 314, 321, 323–325, 339, 340, 343, 352
Mercer’s theorem, 54, 333, 344
method of images, *see* images, method of
microwave engineering, 156, 158, 160
Minakshisundaram-Pleijel zeta function, 88
minimax principle, 283
minimum of a positive definite quadratic form over integer lattice, 67
modular form, 200, 227–231, 233–235, 238, 240, 243, 244, 253, 255, 259, 292–294, 297–301, 303, 304, 311, 313, 315
modular function
  associated with Haar measure, 127
  for the modular group, 240, 265, 289, 292, 302

modular group, $SL(2,\mathbb{Z})$, 197, 203, 205, 214, 216, 219, 230, 238, 242, 259, 268, 272, 289, 314
modular knot, 367
modular symbol, 315
modularity theorem, 300
moduli variety, 242
modulus, λ, 236, 249, 250
moonshine, 242
multiplication formula for Fourier transform, 10, 15, 176
multiplier system, 228, 229, 234, 238, 243–245, 374

N
NaCl, *see* salt
narrow class number, 345, 363
Neumann problem, 195, 273–275, 297
non-Euclidean distance, 152, 153, 193, 209, 334
non-Euclidean Eisenstein series or Epstein zeta function, 339
non-Euclidean geometry, 112, 149, 150, 156
non-Euclidean lattice point or circle problem, 319, 333, 334
non-Euclidean normal distribution, 188, 190, 191
non-Euclidean Poisson summation formula, 333, 334, 340–344
non-Euclidean shock wave, 316
non-Euclidean theta function, 335
norm in field extension, 71, 72, 74, 77, 222, 345
norm of an element of a quaternion algebra, 218–220
norm of an ideal in a number field, 72
norm of hyperbolic element of $SL(2,\mathbb{Z})$, 345
normal density, 27, 29, 186, 188, 190–192, 195
normal or Dirichlet or Poincaré fundamental region, 209, 225
nuclear magnetic resonance tomography, 147

O
$O(y^p)$, 314, 325, 336, 362
$o(y^p)$, 271, 272, 357, 359, 362
orthogonal group or rotation group, $O(n)$ or $SO(n)$, 83, 90, 91, 107, 108, 111, 115, 116, 118, 120, 125, 128, 130–134, 136, 147, 152–155, 178–180, 182, 186, 187, 189, 190, 193, 195, 199, 222, 327, 328, 330, 331, 333, 334, 338, 354
octahedral group, 83
orbital integral, 351, 352, 354, 356
order in an algebra or field, 218–221, 305, 345, 346, 363

P
p-adic number, 60, 82, 131, 189, 303, 375
Paley-Wiener theorem, 16, 168, 175, 178
parabolic element, 317, 319, 344, 345, 351, 355, 360, 362
parabolic fractional linear transformation, 213
Parseval identity, 11, 32
partitions, 238
Pell's equation, 265
Penrose tiling, 97, 98
periodic geodesic, *see* closed geodesic
periodization of test function or distribution, 42
periodogram, 51
perpendicular bisectors method, 208, 209
Peter-Weyl theorem, 130, 131
Petersson inner product, 304, 307
Phragmen–LIndelof theorem, 325
Picard theorem, 202, 241
Plancherel identity, 11, 29
Plancherel measure, 11, 129, 130, 141
Plancherel theorem, 129, 130, 179, 188, 195
Planck constant, 48, 118
Poincaré polygon, *see* normal or Dirichlet or Poincaré fundamental region
Poincaré generators and relations theorem, 204, 211
Poincaré group, *see* fundamental or Poincaré group
Poincaré arc length, 151, 152, 155
Poincaré series, 253–255, 288, 313, 321, 334
Poincaré upper half-plane, H, x, 89, 149, 150, 160, 221, 223
point group, 83
point spectrum, *see* discrete or point spectrum
Poisson integral, 178
Poisson kernel, 178
Poisson summation formula, 39, 40, 42–45, 53, 55, 59, 60, 80, 101, 147, 200, 230, 246, 263, 264, 291, 316, 331–334, 340–344, 356
Poisson's integral formula, 138
polar coordinates, 5, 19, 110, 111, 114, 134, 136, 137, 154, 155, 164, 171–174, 187, 354
positive definite symmetric matrix space, $\mathcal{P}_n$, 261

positive definite symmetric matrix space, $\mathscr{P}_n$, 63, 64, 66, 67, 69, 70, 77, 79, 81, 85, 88, 153–155, 242, 243, 252, 256, 257, 260, 261, 265, 269, 334, 335, 363
positive operator, 54
potential theory, 178, 248, 249
power function, p_s, 165
power function, p_s, 260, 328
power spectrum, 50
prime geodesic theorem, 364
prime number or ideal theorem, 76, 200, 267, 347
primitive hyperbolic element, 346, 347, 353, 362, 364, 367, 374
principal ideal, 72
probability density, 7, 21, 25–29, 121, 142, 188, 190, 191, 193–195
probability distribution, *see* probability density
product of distributions, 6
projection, 97, 373
projection-valued measure, 138
projective linear group, 204, 276, 287, 294, 343, 355, 374
projective plane, 112
projective variety, 256

Q
quadratic form, x, 54, 63, 79, 80, 150, 153, 154, 163, 196, 202, 207, 208, 243, 248, 256, 257, 267, 288, 346, 366
quadratic number field, 73, 75, 83, 207, 241, 273, 284, 346, 347
quadratic reciprocity law, 245
quantum chaos, 365
quantum limit, 281
quantum mechanics, 48, 58, 84, 89, 118, 123, 128, 150, 165, 203, 237, 280, 319, 320
quantum number, 119, 120
quantum-statistical mechanics, 44, 63, 318
quasicrystal, 97, 98, 101
quaternion algebra, 214, 217–219, 221, 317
quiche & salad, 262

R
Rademacher formula for partition function, 238
Rademacher function, 374
Radon transform, R, 142–144, 147, 148
Radon-Nikodym theorem, 25
Ramanujan τ-function, 236, 237
Ramanujan sum, 264
random variable, 7, 25–28, 186, 189–195
Rankin-Selberg method, 312, 313, 322
reciprocity law, 83, 241, 245, 265
regular polyhedra, 147, 223
regulator, 74–77, 79, 81
relative Poincaré series attached to hyperbolic element, 254, 255
representation of a group, *see* group representation
representation of an integer by a quadratic form, 256, 257
residual spectrum, 326
residue of Eisenstein series, 262, 271
residue of zeta function, 64, 76, 80
residue theorem, *see* Cauchy's residue theorem
resolvent, 268, 326, 337, 341, 365
resolvent kernel, *see* Green's function or resolvent kernel
Riemann hypothesis, 66, 67, 81, 93, 262, 283, 284, 305, 314, 317, 325, 360
Riemann mapping theorem, 200
Riemann method of theta functions, 65
Riemann sphere, 240
Riemann surface, 200, 209, 212–214, 218, 219, 221, 334, 347, 364, 366
Riemann zeta function, 61, 63, 66, 75, 81, 82, 200, 229, 244, 256, 257, 261, 265, 274, 295, 298, 302, 360, 362
Riemann–Lebesgue lemma, 32
Riemann-Lebesgue lemma, 12, 13, 143
Riemann-Roch theorem, 232, 235, 366
Riemannian manifold, 89, 110–112, 153, 280, 317, 349
right G-action, 154
right invariant integral, 127, 128, 181, 327, 330
Roelcke-Selberg spectral resolution of Δ on $L^2(\Gamma\backslash H)$, 321, 343
rotation group, *see* orthogonal group or rotation group, $O(n)$ or $SO(n)$
Rydberg constant, 120

S
salt, 83, 86, 87
satellites in spectroscopy, 49
Sato-Tate distribution or Wigner semi-circle distribution, 238
Schrödinger operator, 59, 89, 118–120, 128, 237, 319
Schur's lemma, 126
Schwartz function, 9
Schwarz-Christoffel transformation, 200, 248, 249

second moment, 191, 192
Selberg trace formula, 352, 360, 363, 366, 367, 374
Selberg transform, 178, 329, 344
Selberg zeta function, 365
semidirect product of groups, 136
semisimple Lie group, 128
separation of variables, 36, 114, 115, 164, 167, 173, 185, 186, 263, 338
shah functional, 42
Shannon sampling theorem, 36, 49, 51
Siegel modular form, 245, 247, 343
Siegel modular group, 343
Siegel upper half-space, 245
Siegel zero, 81
singular differential operator, 62, 137, 164
singular eigenvalue problem, 58, 137, 139, 320
singular series, 264
slash operator, 301
Smith chart, 157, 158
smoothing, 34, 35
soliton, 257
Soto-Andrade formula for spherical functions, 225
source spectral density, 49
space group, 83, 147
space of Schwartz functions, $\mathscr{S}$, 9
special linear group, $SL(n)$, 90, 150, 162, 181, 192, 196, 197, 207, 211
$SU(p,q)$, 150, 156, 196
special Lorentz-type group, $SO(p,q)$, 150, 163, 196
special orthogonal group, $SO(n)$, 108
special unitary group, SU(n), 123
spectral lines, 48, 49, 118–121, 128
spectral measure, 138, 141, 168, 169, 176, 321, 327
spectral theorem, 20, 54, 56, 126, 130, 137, 138, 141, 154, 331
spectroscopy, 46, 48, 49
spectrum, 37, 58, 93, 119, 120, 138, 196, 258, 283, 286, 317, 320, 321, 326, 331, 340, 351, 361, 362, 365, 366
sphere, 5, 19, 57, 107, 108
spherical Bessel function, 135
spherical function, 114–116, 132, 172–175, 192, 224, 225, 291, 303, 329, 330
spherical harmonic, 107, 112, 114, 116–119, 121, 123, 125, 129–137, 147, 245
spheroidal wave function, 56
spurious eigenvalues, 273, 274, 341
standard deviation, 26, 28, 192
Stieltjes, Stone, Kodaira, Titchmarsh formula, 137, 139
Sturm-Liouville operator, 113, 137, 139, 140, 164
sun's magnetic field, 121, 123
support of a distribution or function, 4
surface or Laplace spherical harmonic, *see* spherical harmonic
surface spherical harmonic, 112
symmetric power, 163, 237
symmetric space, 63, 89, 108, 112, 130, 136, 150, 173, 175, 223, 224
symplectic group, $Sp(n,\mathbb{R})$, 163, 343

T
table of transforms
  Helgason, 183
  Kontorovich-Lebedev, 170
  Mehler-Fock, 177
  Mellin, 62
Fourier, 17
Tauberian theorem, 24, 28, 45, 267, 314, 337, 340, 362, 364
tautochrone, 24
Tchebychef polynomial, 306
tempered distribution, 15
tensor product of representations, 128
tessellation, 198, 204, 205, 217
test function, 2
tetrahedral group, 83
theta function, 44, 64–66, 88, 200, 228, 242–247, 250, 253, 284, 288, 292, 299, 306
theta group, 200, 212
time-limited function, 55
time-series analysis, 48, 50
topological group, 125
torus, $\mathbb{R}^m/\mathbb{Z}^m$, 41, 101, 102, 129, 196, 214, 215, 250, 317, 347
torus graph, 91
totally real number field, 71, 217, 221, 304, 305, 365
trace formula, 352, 360–363, 365–367, 374, 375
trace of an element of an extension field, 71, 92, 289
trace of an operator, 54, 55, 129
transcendental number, 82, 242
transformation formula for
  a holomorphic modular form of weight k, 293
  eta, 238, 374
  theta, 44, 64, 88, 243, 245, 246
  theta differentiated, 66
transmission line, 157, 158, 165, 189, 195

trefoil knot, 367
twisted L-function, 301
twisted trace formula, 300, 317

U
uncertainty principle, *see* Heisenberg uncertainty principle
uniform distribution, *see* distribution
unimodular group, 128, 129
unique factorization, 70, 72, 73, 75
unit disc, 150, 156, 178, 198, 205, 217
unit group in number field, 73, 74, 83
unitary representation, 20, 125–130, 224, 374, 375
universal covering surface, 200
unramified field extension, 241

V
value of zeta or L-function, 64, 74, 79, 82, 83
variance, 26, 191, 192
vector space of holomorphic cusp forms of weight k, $\mathscr{S}(\Gamma,k)$, 231
vector space of holomorphic modular forms of weight k, $\mathscr{M}(SL(2,\mathbb{Z}),k)$, 228, 233, 234, 240, 243, 254, 257, 269, 292–294, 301, 303, 304
vector space of Maass cusp forms, $\mathscr{S}\mathscr{N}(SL(2,\mathbb{Z}),\lambda)$, 268, 272, 285, 287, 314, 360, 362
vector space of Maass waveforms, $\mathscr{N}(SL(2,\mathbb{Z}),\lambda)$, 259, 261, 262, 268–270, 295–298, 306, 307, 309
venus spectra, 46
vibrating
  drum, 36, 37, 39, 366
  manifold, 366
  plate, 39
  rod, 31
  string, 14
voltage reflection coefficient, 157
volume of fundamental domain for $SL(3,\mathbb{Z})$, 265
von Neumann spectral theorem, 137, 138

W
wave equation, 14, 20, 36, 38, 60, 117, 147, 196, 317
wave number, 48
wavelets, 94, 96, 97
Weierstrass continuous nowhere differentiable function, 105, 106
Weierstrass function, $\wp$, 235, 236, 247, 248, 250, 258
weight enumerator of a code, 253
weight for a space of functions, 139
weight of a holomorphic modular form, 200, 227–233, 240, 255, 256
weight of a representation, 129
Weil estimate for Kloosterman sums, 93
Weil explicit formulas, 360
Weil-Hecke theory, 299, 300
Weyl character formula, 128
Weyl ergodic theorem or criterion for uniform distribution, 36, 101, 102, 347
Weyl's theorem on asymptotic distribution of eigenvalues of Laplacian on a compact domain, 45, 360
Wiener expression for Fourier transform on $\mathbb{R}$, 59
Wiener integral, 318
Wiener–Khintchine formula, 50
Wigner semi-circle distribution, 237

Z
Zeeman effect, 120
zero of doubly periodic function, 246
zeros of holomorphic modular forms, 232
zeros of zeta and L-functions, 66, 67, 81, 262, 273, 305, 341, 360
zeta function
  Dedekind, *see* Dedekind zeta function
  Epstein, *see* Epstein zeta function
  from a modular form, 300
  Ihara, 365
  Riemann, *see* Riemann zeta function
  Ruelle and Smale, 365
  Selberg, *see* Selberg zeta function
zonal spherical function, 115